



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

Inversion start site

```
860 |
ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
+-----+-----+-----+-----+ 900
TAATATTTCTTTTTCTTTTATTGCGTTACCTGTTCAACAC (41)
Y K G K R K * R N G Q V V

AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAACCAAAGTGAGGTAGA
901 +-----+-----+-----+-----+ 960
TTGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT (101)
K L * T Q V C T I I R N T P K P K * G R

AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961 +-----+-----+-----+-----+ 996
TTATCGTACTCTTCGGCACAACTACAATTAATTAA (137)
N S M R S R V * C * L I
```

The inversion sequence of the apo-dystrophin-4 cDNA (SEQ ID NO: 1)

Figure 1



RECEIVED
SEP 18 2002
TECH CENTER 1600/2900

Inversion start site
|
TAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
850 +-----+-----+-----+-----+-----+ 900
ATTTCTTTCTTAATATTCCTTTTCTTTTATTGCGTTACCTGTTCAACCAC (51)
* R K N Y K G K R K * R N G Q V V

AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAACCAAAGTGAGGTAGA
901 -----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT (111)
K L * T Q V C T I I R N T P K P K * G R

AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961 -----+-----+-----+-----+-----+ 996
TTATCGTACTCTTCGGCACAACTACAATTAATTAA (147)
N S M R S R V * C * L I

The inversion sequence of the apo-dystrophin-4 cDNA plus a 10 base-pair region 5' to the start of the inversion sequence (SEQ ID NO: 1A, included in SEQ ID NO: 2).

Figure 1A



Start at 710

|
AACAATGGCAG
+-----+ 720
TTGTTACCGTC (11)
Q W Q

```
721  GTTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAACCTACATGTAAAATCTTG 780
      -----+-----+-----+-----+-----+-----+-----+
      CAAAATGTGCAGATACGTTAACATGTTTTTCAATATTCTTTGATGTACATTTTAGAAC (71)
      V L H V Y A I V Q K S Y K K T T C K I L

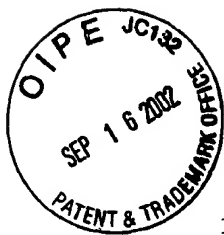
781  ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA 840
      -----+-----+-----+-----+-----+-----+-----+
      TATCGATTTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTAAAT (131)
      I A K * L A I S L Y G T H F G L F K N L
              inversion start site
              |
841  TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
      -----+-----+-----+-----+-----+-----+-----+
      ATTGTCAATATTTCTTTCTTAATATTTCTTTTCTTTTATTGCGTTACCTGTTCCACCAC (191)
      * Q L * R K N Y K G K R K * R N G Q V V

901  AAGCTGTGAACCTCAGGTGTGCACAATTATCAGGAACACCCCAAACCAAAGTGAGGTAGA 960
      -----+-----+-----+-----+-----+-----+-----+
      TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT (251)
      K L * T Q V C T I I R N T P K P K * G R

961  AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996
      -----+-----+-----+-----+-----+
      TTATCGTACTCTTCGGCACAAACTACAATTAATTAA (287)
      N S M R S R V * C * L I
```

The inversion sequence of the apo-dystrophin-4 cDNA plus the upstream 150 bp from the start of the inversion at 860 to the Hpa I enzyme site (SEQ ID NO: 1B, included in SEQ ID NO: 2)

Figure 1B



GTGGTTTGATTGATAGTAAAAAATGTTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT
1 -----+-----+-----+-----+-----+-----+ 60
CACCAAATACTATCATTTTTTTTTACAAGCAATTATGTTTCATCTCTCATTTCATTAGTTA
V V * L I V K K M F V N T S R E * V I N
CAATCACTCATAGCCAAGGTGAAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
61 -----+-----+-----+-----+-----+-----+ 120
GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
Q S L I A K V E K M Y P I M E Y S C S D
AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
121 -----+-----+-----+-----+-----+-----+ 180
TCTTTAGAACACGAATAGATACCTTAAGAAAATATATATAAATGTAACCCTTGGAAGTTA
R N L V L I Y G I L L I Y I Y I G N L N
GTAGCTTGACATTTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
181 -----+-----+-----+-----+-----+-----+ 240
CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
V A * H F S M * T P V A * S N I K L I L
ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAACTGGTGAAA
241 -----+-----+-----+-----+-----+-----+ 300
TGTTTGTTGCACATTACCGAAGTAATTATTCGAAACGAAGAAGGACCTTTGACCACTTT
T N N V * W L H * * G F A S S W K L V K
AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
301 -----+-----+-----+-----+-----+-----+ 360
TTAGTTTGGAACAACACATGTGGGAGCTACGTCGAAGACACAACAGAAGTGGGTCTTTAC
N Q T L L C T P S M Q L L C C L H P E M

The polynucleotide sequence of apo-dystrophin-4 (SEQ ID NO: 2)

Figure 2



361 GGG AATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA
-----+-----+-----+-----+-----+-----+
420 CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
G N D F P N G K E T E * C Y L S A P F V

421 AAGTCTGTCTTTCTTTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG
-----+-----+-----+-----+-----+-----+
480 TTCAGACAGAAAGAAAGAGAAACAAAAGGTCCTGTGTACATCCTTCAGAAAAGGTGTAC
K S V F L S L C F P G H N V G S L F H M

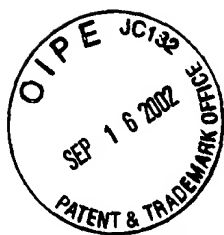
481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
-----+-----+-----+-----+-----+-----+
540 CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT
A D D L G R A M E S L V S V M T D E E G

541 GCAGAATAAATGTTTTACAACCTCCTGATTCCTCGCATGGTTTTTATAATATTCATACAACA
-----+-----+-----+-----+-----+-----+
600 CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT
A E * M F Y N S * F P H G F Y N I H T T

601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG
-----+-----+-----+-----+-----+-----+
660 TTCTCCTAATCTGTCATTCTCAAATGTTCTTTATTTAGATATAAAAACACTTCCCATCAC
K R I R Q * E F T R N K S I F L * R V V

661 GTATTATACTGTAGATTTCACTAGTTTCTAAGTCTGTTATTGTTTTGTAAACAATGGCAG
-----+-----+-----+-----+-----+-----+
720 CATAATATGACATCTAAAGTCATCAAAGATTTCAGACAATAACAAAACAATTGTTACCGTC
V L Y C R F Q * F L S L L L F C * Q W Q

Figure 2 (cont'd)



721 GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAACTACATGTAAAATCTTG
-----+-----+-----+-----+-----+-----+ 780
CAAAATGTGCAGATACGTTAACATGTTTTTCAATATTCTTTGATGTACATTTTAGAAC
V L H V Y A I V Q K S Y K K T T C K I L

781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTGTTTAAAAATTTA
-----+-----+-----+-----+-----+-----+ 840
TATCGATTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTAAAT
I A K * L A I S L Y G T H F G L F K N L

841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
-----+-----+-----+-----+-----+-----+ 900
ATTGTCAATATTTCTTTCTTAATATTTCTTTTTCTTTTATTGCGTTACCTGTTCAACCAC
* Q L * R K N Y K G K R K * R N G Q V V

901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
-----+-----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT
K L * T Q V C T I I R N T P K P K * G R

961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
-----+-----+-----+-----+-----+ 996
TTATCGTACTCTTCGGCACAACTACAATTAATTAA
N S M R S R V * C * L I

Figure 2 (cont'd)

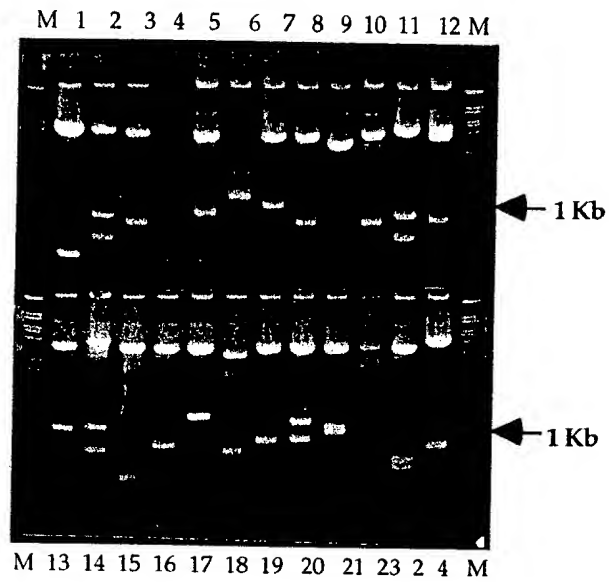


Figure 3A

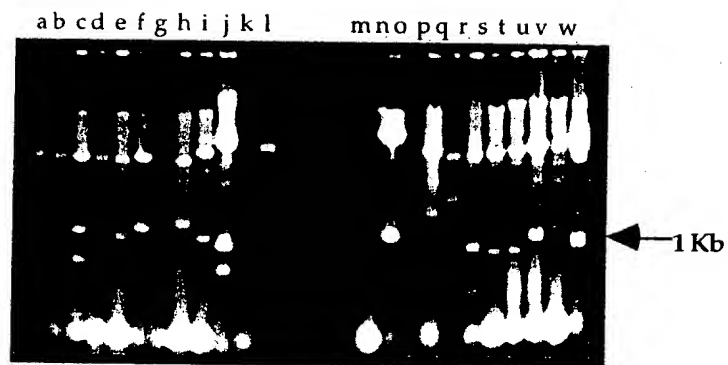


Figure 3B



SEP 18 2002

TECH CENTER 1600/2900

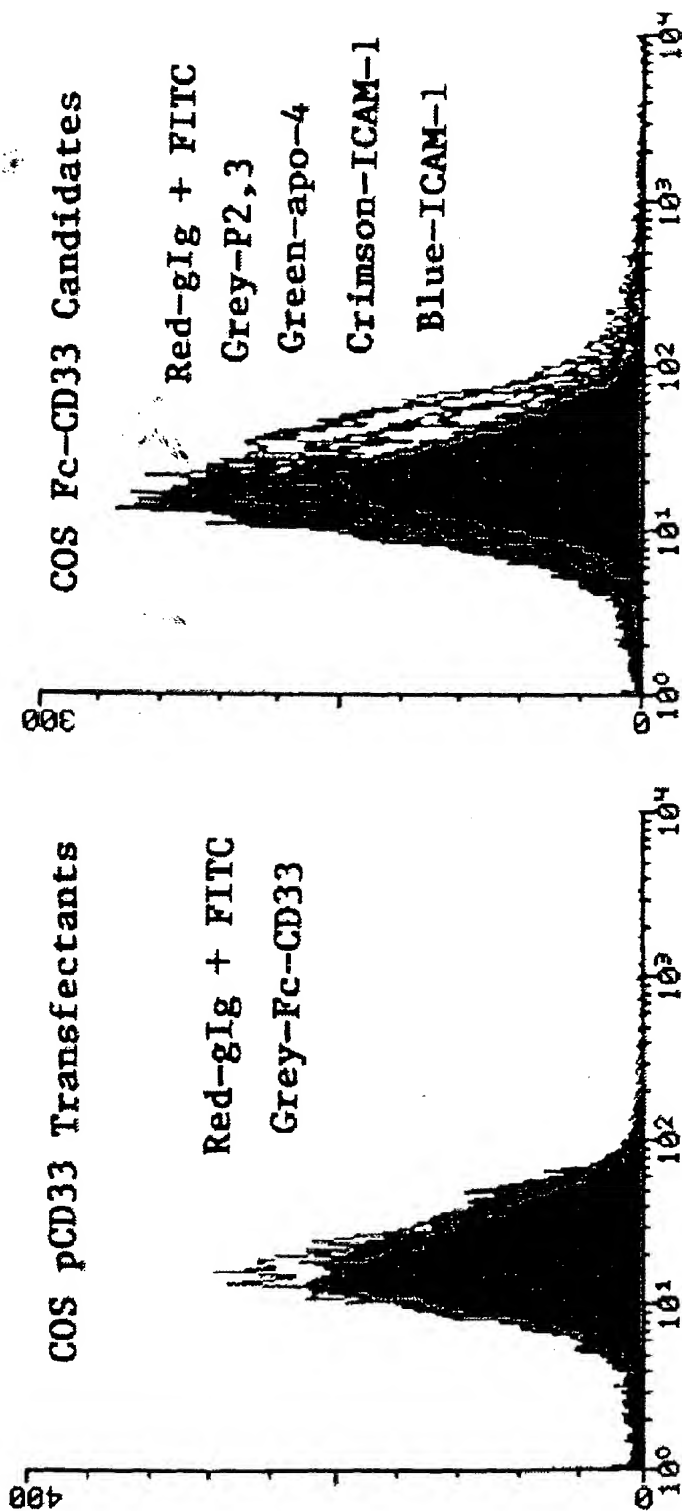


Figure 4B

Figure 4A

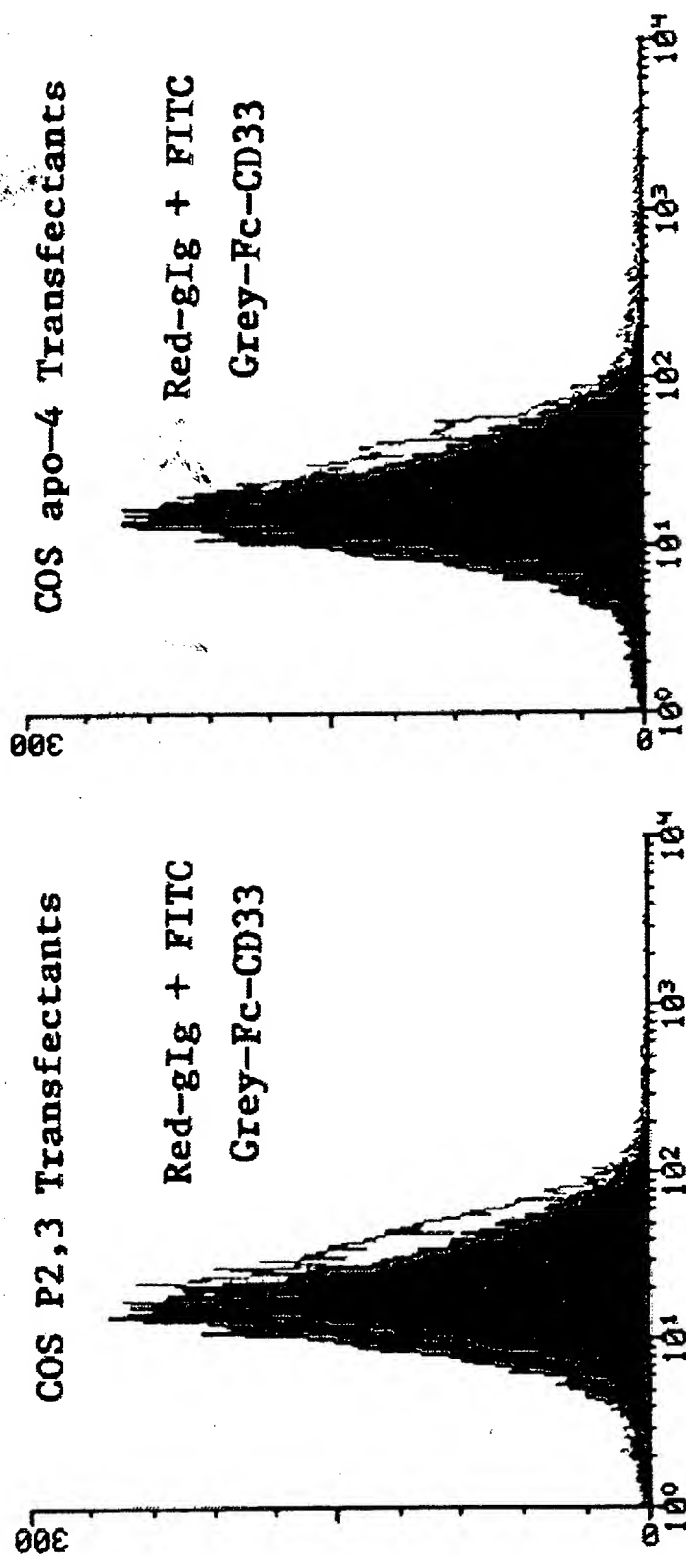


Figure 4D

Figure 4C

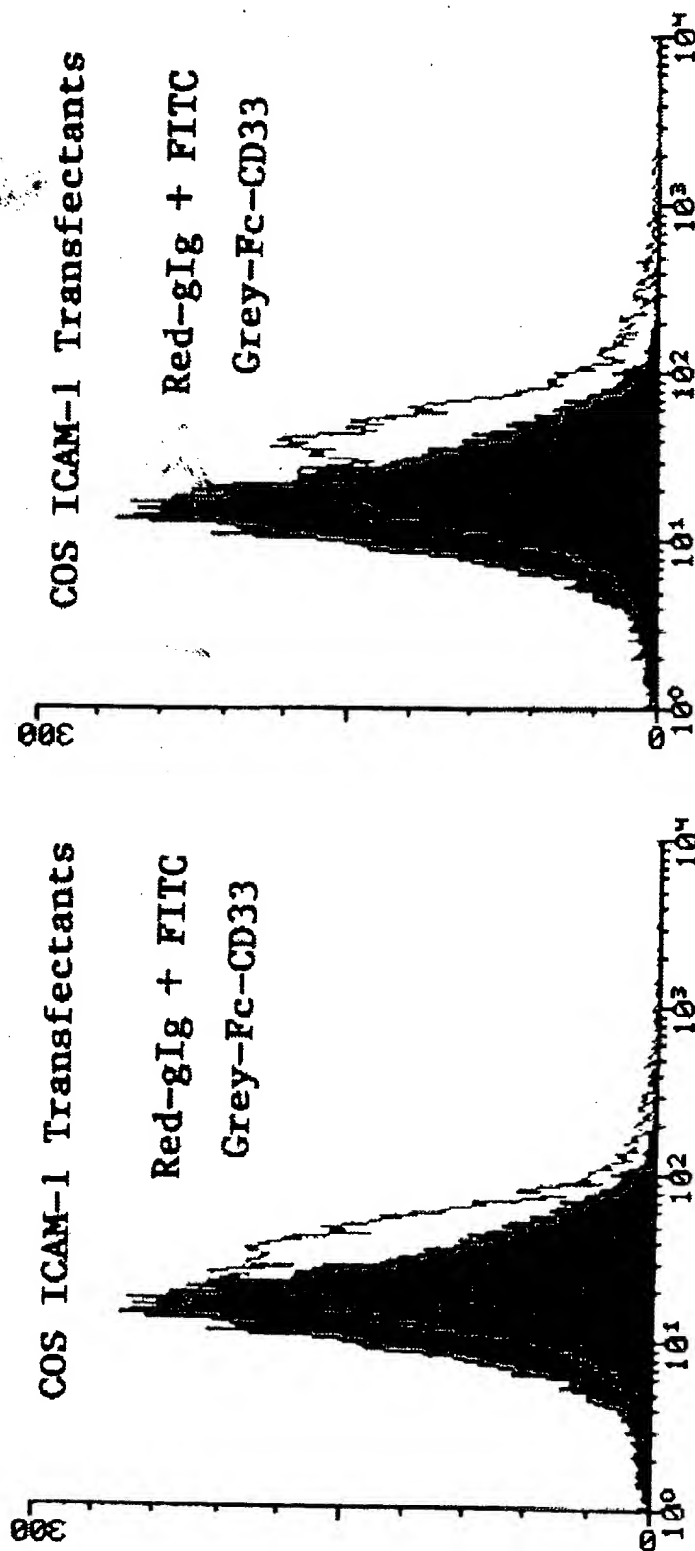


Figure 4F

Figure 4E

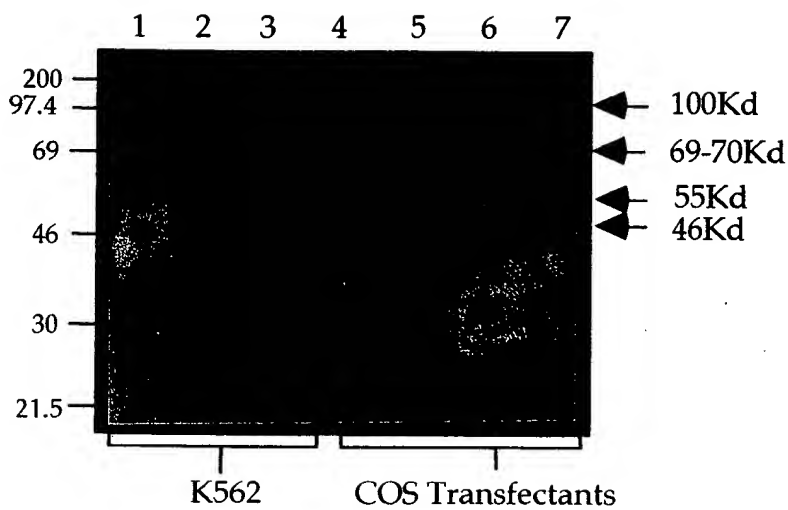


Figure 5

. . TAGTTTCCTATTCTCAATGTATAGTAGCCACCAAGAGGTCAATTCAGAGGTTTATTATTATT
 -239 -180
 . . . ATCAAAGGATAAGTTACATATCACGTGGTTTCCAGTTAAGTTCTCAAATAATAATAA
 . * F P I Q C I V H Q R S I Q E F I I I
 ATTTTCAACCCAAGTAAAAGCAGAGAGAAAATAGCCACCTCCACCATAGCCTCAGAAGCA
 -179 -120
 TAAAAGTTGGGTTCATTTTCGTCTCTCTTTTATCGGTGGAGGTGGTATCGGAGTCTTCGT
 I F N P S K S R E K I A T S T I A S E A
 AGCCAACAGCCTGAAACAGCTTTGAAATGAAAAGTTGGTGTGGCGGTGATGGTGGCAGTG
 -119 -60
 TCGGTTGTGCGACTTTGTGCGAACTTTACTTTTCAACCACACCGCCACTACCACCGTCAC
 S Q Q P E T A L K * K V G V A V M V A V
 ATAATGGTGACCGATGGTTGGGTGCTGGTGATGGTAGTGGTAGTTGTGAAGGTGGTGATG
 -59 0
 TATTACCACTGGCTACCAACCCACGACCACTACCATCACCATCAACACTTCCACCACTAC
 I M V T D G W V L V M V V V V V V K V V M
 GTGGTTTGATTGATAGTAAAAAAATGTTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT
 1 60
 CACCAAACCTAACTATCATTTTTTTTACAAGCAATTATGTTTCATCTCTCATTATTAGTTA
 V V * L I V K K M F V N T S R E * V I N
 CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
 61 120
 GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
 Q S L I A K V E K M Y P I M E Y S C S D
 AGAAATCTTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
 121 180
 TCTTTAGAACCGAATAGATACCTTAAGAAACTATATATAAATGTAACCTTGGACTTA
 R N L V L I Y G I L L I Y I Y I G N L N
 GTAGCTTGACATTTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
 181 240
 CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
 V A * H F S M * T P V A * S N I K L I L
 ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA
 241 300
 TGTTTGTGTCACATTACCGAAGTAATTATTCGAAACGAAGAAGGACCTTTGACCACTTT
 T N N V * W L H * * G F A S S W K L V K
 AATCAAACCTTGTTGTGTACACCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
 301 360
 TTAGTTTGGAAACAACACATGTGGGAGCTACGTGGAAGACACAACAGAAGTGGGTCTTTAC
 N Q T L L C T P S M Q L L C C L H P E M

Figure 6 (SEQ ID NO: 51)



361 GGG AATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA 420
-----+-----+-----+-----+-----+-----+
CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
G N D F P N G K E T E * C Y L S A P F V
begin exon 79
|
421 AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480
-----+-----+-----+-----+-----+-----+
TTCAGACAGAAAGAAAGAGAAACAAAAGGTCCTGTGTACATCCTTCAGAAAAGGTGTAC
K S V F L S L C F P G H N V G S L F H M
GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
481 -----+-----+-----+-----+-----+-----+ 540
CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT
A D D L G R A M E S L V S V M T D E E G
GCAGAATAAATGTTTTACAACCTCGATTCCCGCATGGTTTTTATAATATTCATACAACA
541 -----+-----+-----+-----+-----+-----+ 600
CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT
A E * M F Y N S * F P H G F Y N I H T T
(---N---)
601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660
-----+-----+-----+-----+-----+-----+
TTCTCCTAATCTGTCAATCTCAAATGTTCTTTATTTAGATATAAAAACTTCCCATCAC
K R I R Q * E F T R N K S I F L * R V V
GTATTATACTGTAGATTTCAGTAGTTTCTAAGTCTGTTATTGTTTGTGAACAATGGCAG
661 -----+-----+-----+-----+-----+-----+ 720
CATAATATGACATCTAAAGTCATCAAAGATTGAGACAATAACAAAACAATTGTTACCGTC
V L Y C R F Q * F L S L L L F C * Q W Q
GTTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAACCTACATGTAAAATCTTG
721 -----+-----+-----+-----+-----+-----+ 780
CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC
V L H V Y A I V Q K S Y K K T T C K I L
ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTGTTTAAAAATTTA
781 -----+-----+-----+-----+-----+-----+ 840
TATCGATTTATTGAACGGTAAAGAAATATACCTTGCCTAAAACCCAACAAATTTTAAAT
I A K * L A I S L Y G T H F G L F K N L
inversion start site
|
841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900
-----+-----+-----+-----+-----+-----+
ATTGTCAATATTTCTTTCTTAATATTTCTTTTCTTTTATTGCGTTACCTGTTCCACCAC
* Q L * R K N Y K G K R K * R N G Q V V

Figure 6 (cont'd)



901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAACCAAAGTGAGGTAGA 960
-----+-----+-----+-----+-----+-----+
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT
K L * T Q V C T I I R N T P K P K * G R

AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961 -----+-----+-----+-----+-----+ 996
TTATCGTACTCTTCGGCACAAACTACAATTAATTAA
N S M R S R V * C * L I

Figure 6 (cont'd)

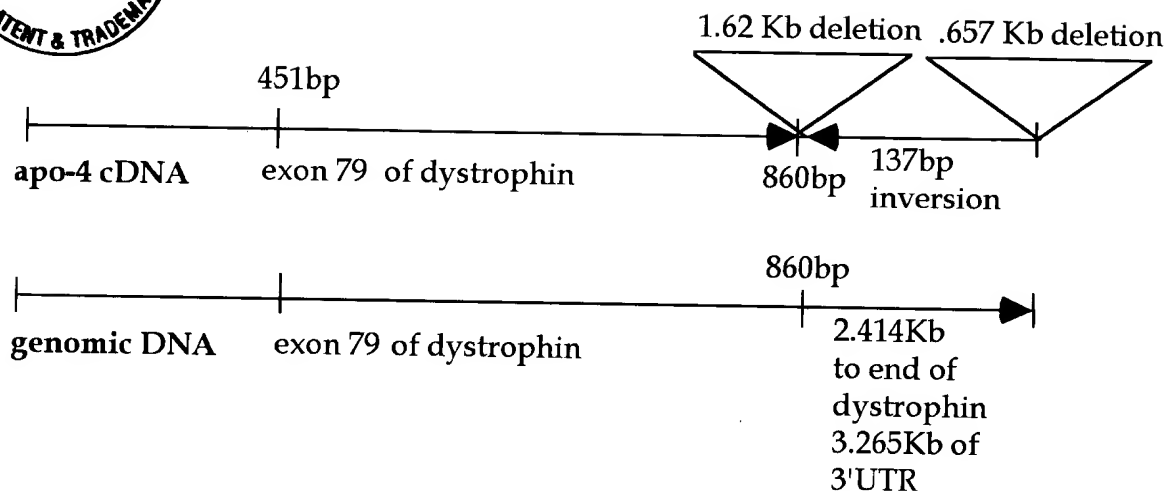
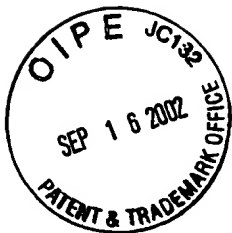
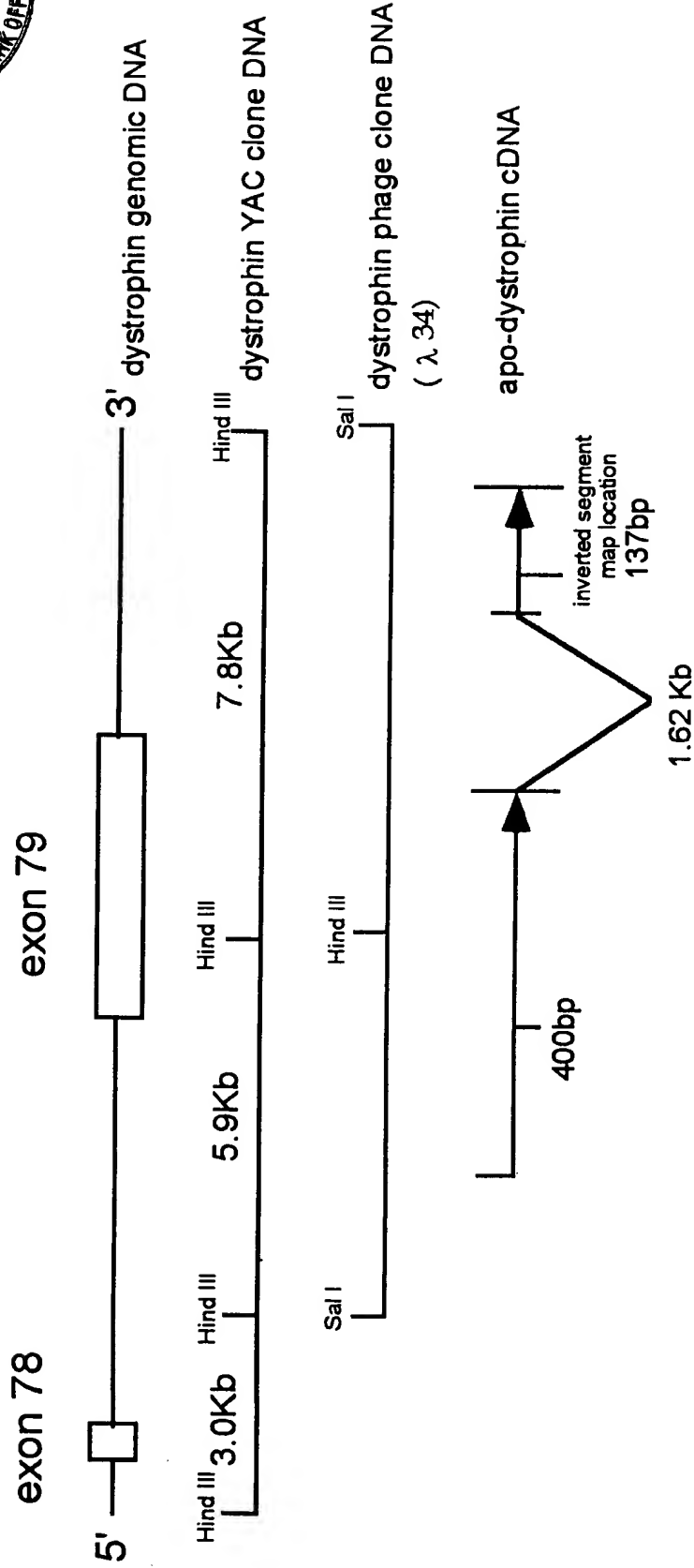
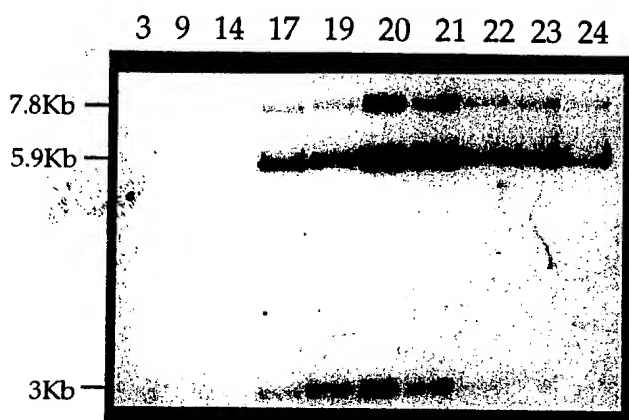


Figure 7



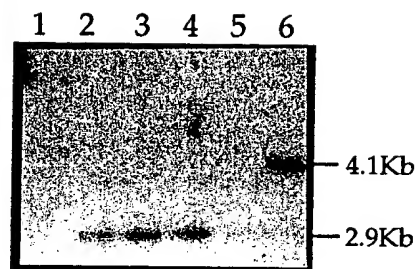
*cDNA map is not precisely drawn to scale

Figure 8



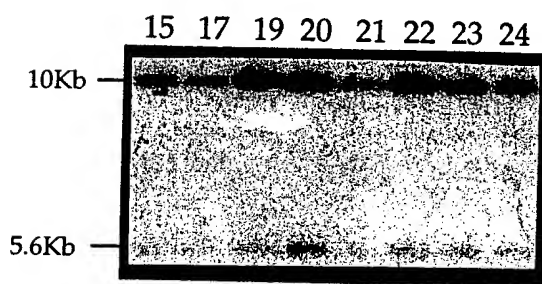
A. Human

Figure 9A



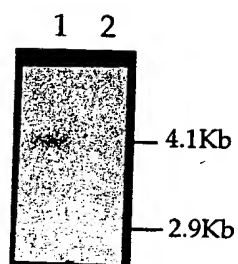
B. Mouse

Figure 9B



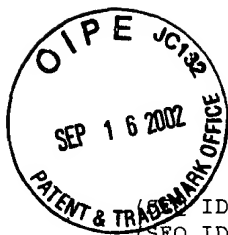
A. Human

Figure 10A



B. Mouse

Figure 10B



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

```

1
ID NO 54) Mgen1073 ..... 50
(SEQ ID NO 55) Hap01234 ctagtttcct attcaatgta tagtgcacca aaggtcaatt caagagttta
Consensus -----

51
Mgen1073 ..... 100
Hap01234 ttattattat ttccaaccca agtaaaagca gagagaaaat agccacctcc
Consensus -----

101
Mgen1073 ..... begin GRAIL exon @149 150
Hap01234 accatagcct cagaagcaag ccaACAGcCT gAAaCAGCtt tgAAATGAaA
Consensus -----ACAG-CT -AA-CAGC-- --AAATGA-A

151
Mgen1073 AtT.....T AtgtGgtAgt cAgGtcactG
Hap01234 AgTtggtgtg gcggtgatgg tggcagtgaT AatgGtgAcc gAtGgttggG
Consensus A-T-----T A---G--A-- -A-G-----G

201
Mgen1073 TGCTGGTaAT GGTgaTctTA GcaGgcAgAG aaGGTGgTaG TGaTTTGATa
Hap01234 TGCTGGTgAT GGTagTggTA GttGtgA.AG gtGGTGaTgG TGgTTTGATt
Consensus TGCTGGT-AT GGT--T--TA G--G--A-AG --GGTG-T-G TG-TTTGAT-

251
Mgen1073 GtaAaagtgt AgAcTaTaCa acAgaAtAAa TACaagtatA GTAA.....
Hap01234 GatAgtaaaa AaAaTgTtCg ttAatAcAAg TAgAgagtaA GTAAatcaatc
Consensus G--A-----A-A-T-T-C- --A--A-AA- TA-A-----A GTAA-----

301
Mgen1073 .....atc caaCAAAGTG tgAAAGgTGT gTgCCATtAc acAtctTTCt
Hap01234 aatcactcat agcCAAAGTG gaAAAGaTGT aTcCCATcAt ggAataTTCc
Consensus -----CAA-GTG --AAAG-TGT -T-CCAT-A- --A---TTC-

351
Mgen1073 cG.....GtgATAagag cCTTgTCTAT GaAgTTC... TGAgATgTgT
Hap01234 tGttctgata GaaATcttgt gCTTaTCTAT GgAaTTCttt TGAtATaTaT
Consensus -G-----G--AT-----CTT-TCTAT G-A-TTC--- TGA-AT-T-T

401
Mgen1073 TaggAagatG AATcatcAat TtaCaT... TTcTcCCcat cAAAtgaCAC
Hap01234 TtacAttggG AAACtgaAtg TagCtTgaca TTtTtCCatg tAAAcacCag
Consensus T--A---G AA-C---A-- T--C-T--- TT-T-CC--- -AAA---CA-

451
Mgen1073 ..... begin mouse GRAIL exon ..... 500
Hap01234 cAtgCTGATC CAGtATTAAG CTaATACTAA C.....ACca tgcAatGCTT
Consensus -A--CTGATC CA--ATTAAG CT-ATACTAA C-----AC-- ---A--GCTT

501
Mgen1073 CATTAACaAG GaTTTGCTTC TTgCTaGAAA tgGGT..AAA AaCggActgT
Hap01234 CATTAAtAAG GcTTTGCTTC TTcCTgGAAA ctGGTgaAAA AtCaaACctT
Consensus CATTA-AAG G-TTTGCTTC TT-CT-GAAA --GGT--AAA A-C--AC--T
```

Figure 11



-70 bp from 5' end of apo-4

|

Inr = GCCC TCAT TCTG GAGAC (SEQ ID NO: 59)

apo-4 = GCGG TGAT GGTG GCAGT (SEQ ID NO: 60)

- 48% perfect homology with Inr

71% match on type of base

(purine vs. pyrimidine)

Figure 12B



	551		600
Mgen1073	GgTcTGTAtA CcTtCaATGC AGCTTaTGTG TTGTCTTttc C..tgAAatG		
Hapo1234	GtTgTGTAcA CCcTCgATGC AGCTTcTGTG TTGTCTTcaC CcagaAAatgG		
Consensus	G-T-TGTa-A CC-TC-ATGC AGCTT-TGTG TTGTCTT--C C----AA--G		
	601		650
Mgen1073	GtAATGAcTc CCaAtAgtGg cAAccAgggG tacaATaCT.TGCA		
Hapo1234	GgAATGAtTt CCcAaAtgGc aAAgaAacaG agtgATgCTa tctatcTGCA		
Consensus	G-AATGA-T- CC-A-A--G- -AA--A---G ----AT-CT- -----TGCA		
	651		exon79 700
Mgen1073	CacTTTGTA A...cTCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT		
Hapo1234	CctTTTGTA AgtctgTCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT		
Consensus	C--TTTGTA A-----TCTT TCTTTCTCTT TGTTTTCCAG GACACAATGT		
	701		750
Mgen1073	AGGAAGcCTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT		
Hapo1234	AGGAAGtCTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT		
Consensus	AGGAAG-CTT TTCCACATGG CAGATGATTT GGGCAGAGCG ATGGAGTCCT		
	751		800
Mgen1073	TAGTtTCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC		
Hapo1234	TAGTaTCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC		
Consensus	TAGT-TCAGT CATGACAGAT GAAGAAGGAG CAGAATAAAT GTTTTACAAC		
	801		850
Mgen1073	TCCTGATTCC CGCATGGTTT TTATAATATT CgTACAACAA AGAGGATTAG		
Hapo1234	TCCTGATTCC CGCATGGTTT TTATAATATT CaTACAACAA AGAGGATTAG		
Consensus	TCCTGATTCC CGCATGGTTT TTATAATATT C-TACAACAA AGAGGATTAG		
	851		900
Mgen1073	ACAGTAAGAG TTTACAAGAA ATaAAATCTA TATTTTTGTG AAGGGTAGTG		
Hapo1234	ACAGTAAGAG TTTACAAGAA AT.AAATCTA TATTTTTGTG AAGGGTAGTG		
Consensus	ACAGTAAGAG TTTACAAGAA AT-AAATCTA TATTTTTGTG AAGGGTAGTG		
	901		950
Mgen1073	GTAcTATACT GTAGATTTCa GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA		
Hapo1234	GTAtTATACT GTAGATTTCa GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA		
Consensus	GTA-TATACT GTAGATTTCa GTAGTTTCTA AGTCTGTTAT TGTTTTGTTA		
	951		1000
Mgen1073	ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAA AGTTAaAAGA		
Hapo1234	ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAA AGTTAtAAGA		
Consensus	ACAATGGCAG GTTTTACACG TCTATGCAAT TGTACAAAA AGTTA-AAGA		
	1001		1050
Mgen1073	AA...ACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT		
Hapo1234	AAactACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT		
Consensus	AA---ACATG TAAAATCTTG ATAGCTAAAT AACTTGCCAT TTCTTTATAT		

Figure 11 (cont'd)



```

begin inversion@1100
1051
Mgen1073 GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGAt
Hap1234 GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGAA
Consensus GGAACGCATT TTGGGTTGTT TAAAAATTTA TAACAGTTAT AAAGAAAGA-

1101
Mgen1073 TgtAAActaA Agtgtgcttt AtAAAAaAAG ttgtTtataA AaaccctAa
Hap1234 TtatAAaggA A.....aa AgAAAAAtAAc gcaaTggacA AgtgggtgaAg
Consensus T---AA---A A----- A-AAAA-AA- ----T----A A-----A-

1151
Mgen1073 acaaacACaC AcGcacaCAC AcacAcacac AcacaCaCaC AcaCaCaAcTG
Hap1234 ctgtgaACtC AgGtgtgCAC AattAtcagg AacacCcCAa AacCAaAgTG
Consensus -----AC-C A-G----CAC A---A----- A----C-CA- A--CA-A-TG

1201
Mgen1073 AGGcAGcAca ttgtTttGcA ttacTtTagc gTGTatcaTA t..
Hap1234 AGGTAGaAat agcaTgaGaA gccgTgTttg aTGTtaatTA att
Consensus AGG-AG-A-- ----T--G-A ----T-T--- -TGT----TA ---

```

Figure 11 (cont'd)

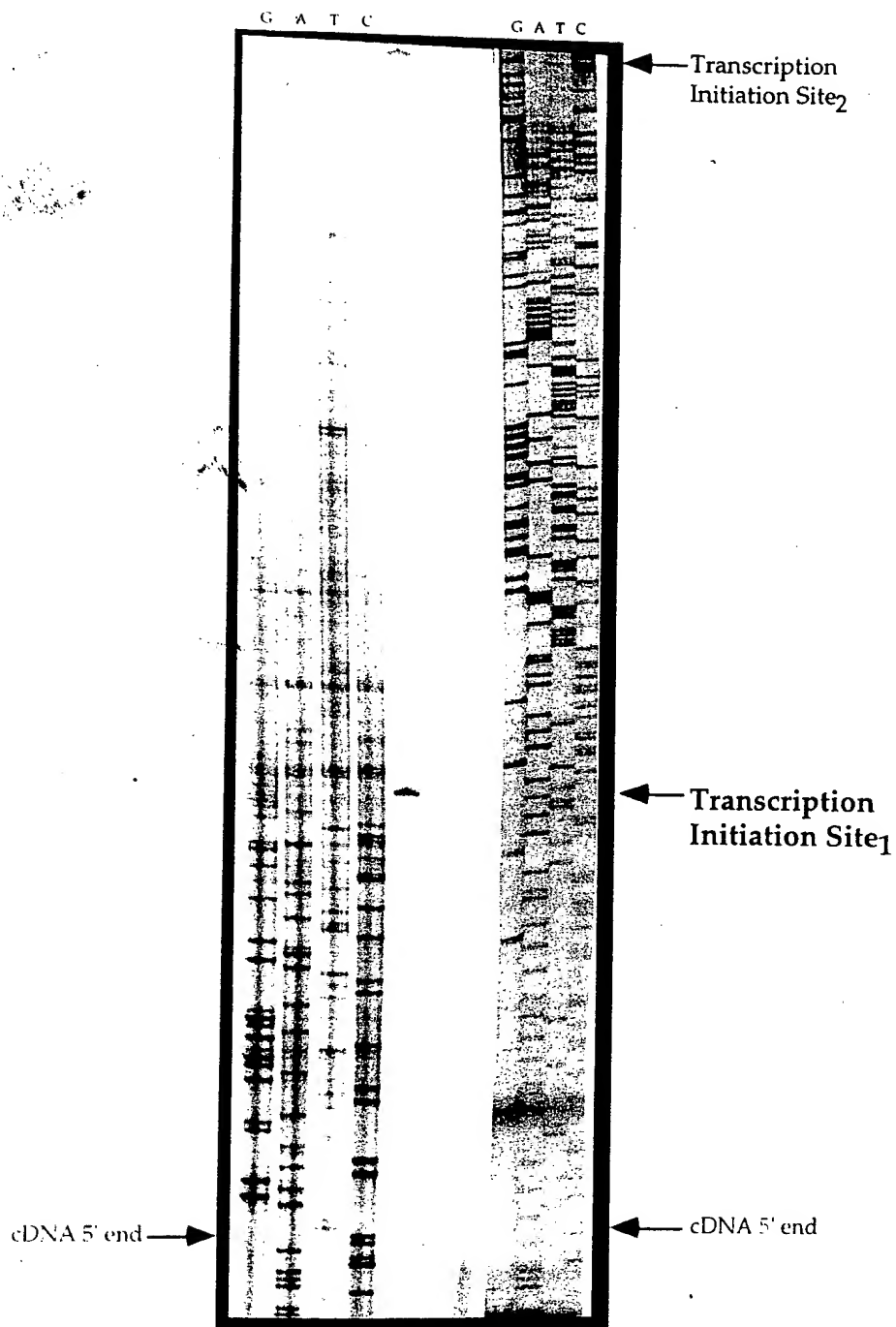


Figure 12A

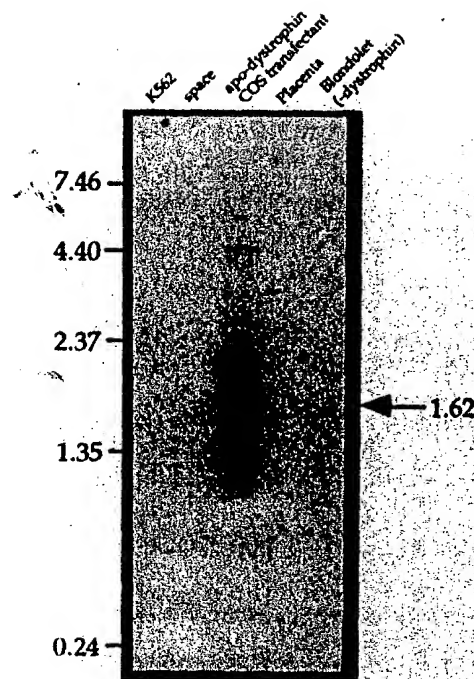


Figure 13

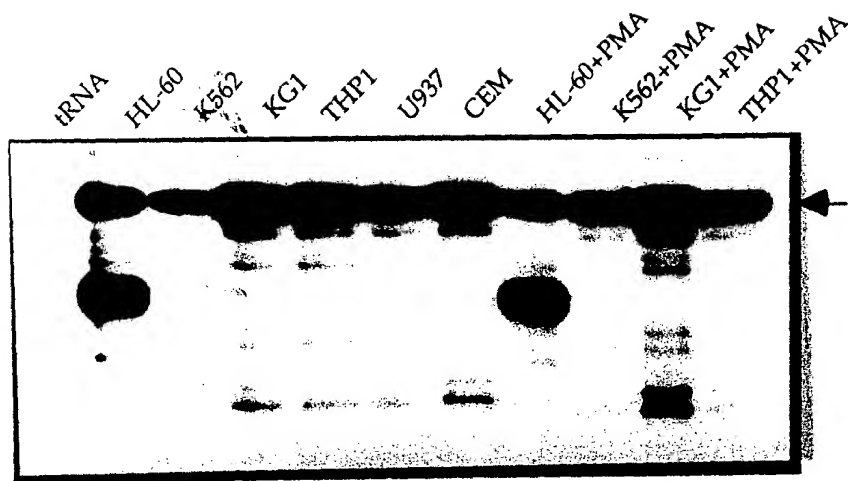


Figure 14

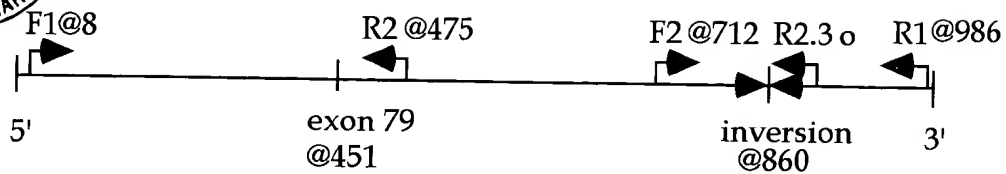
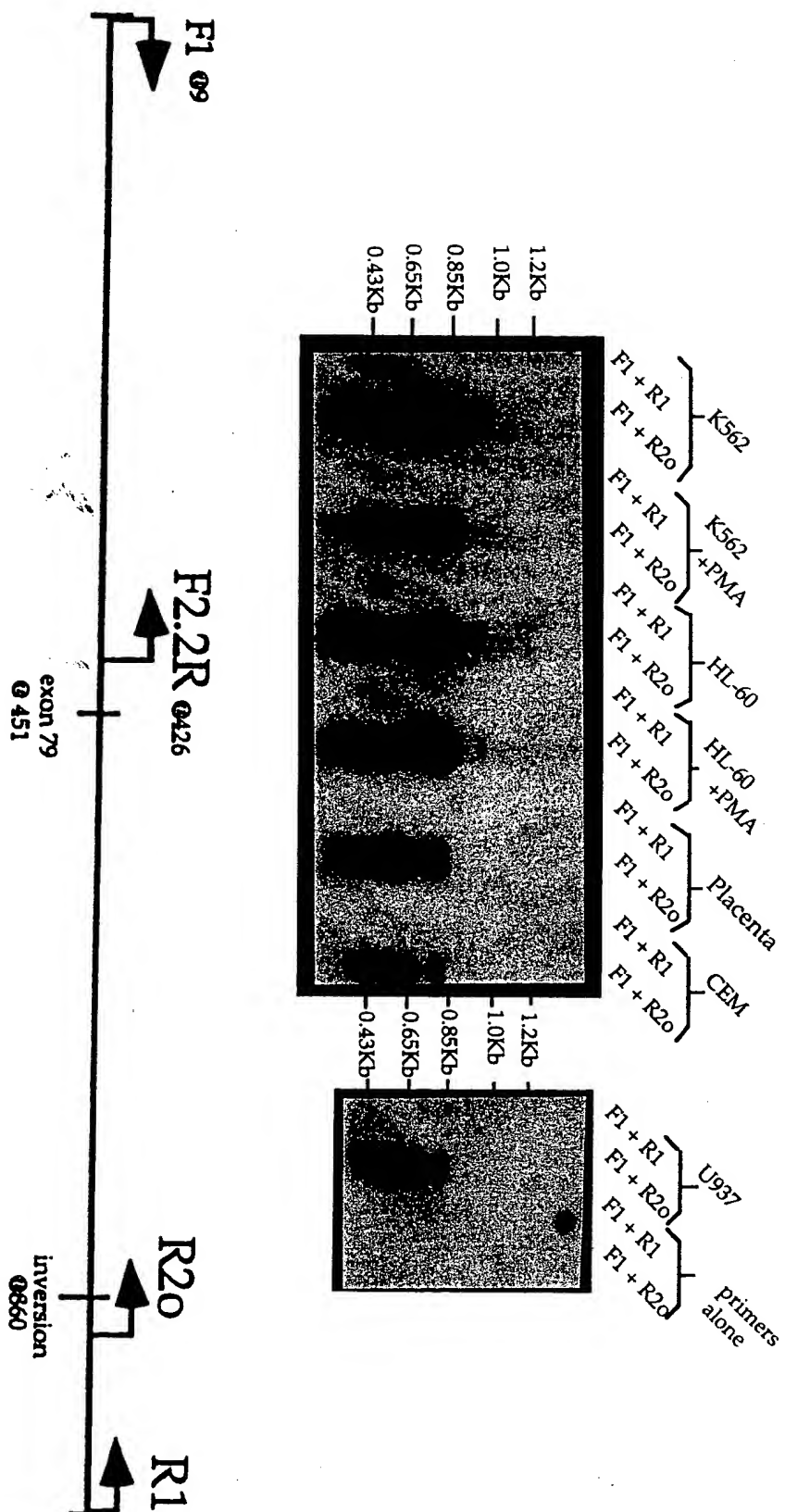


Figure 15

The Apo-dystrophin cDNA

Figure 16



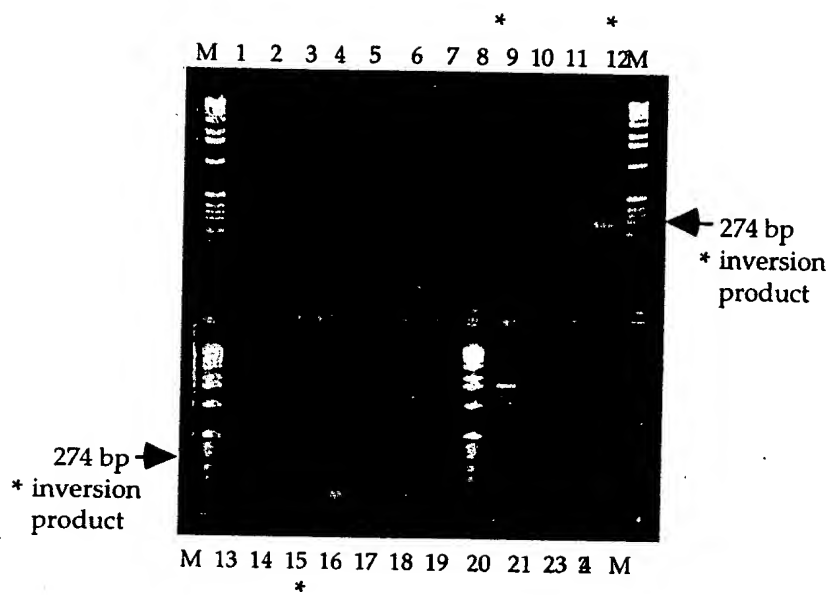
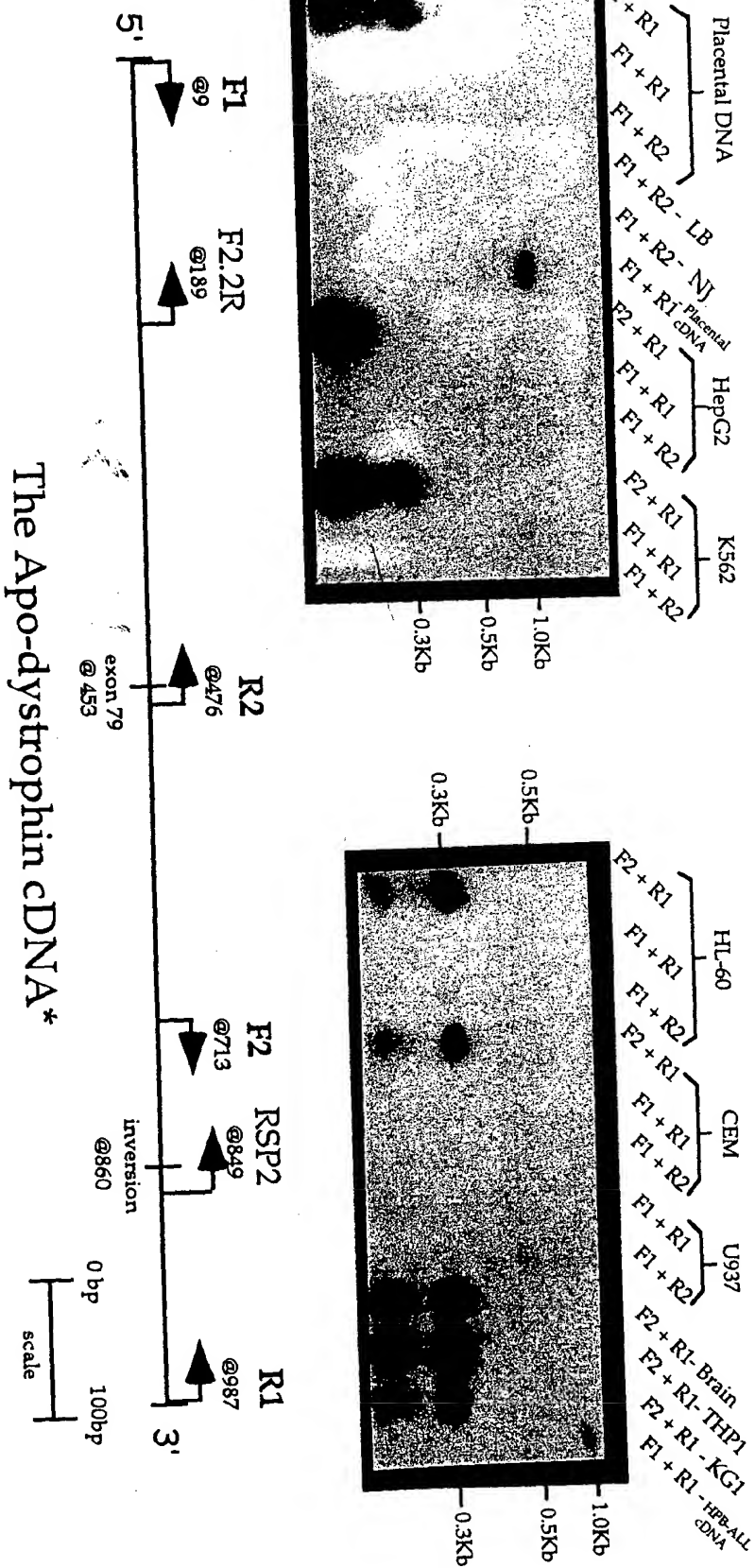


Figure 17A

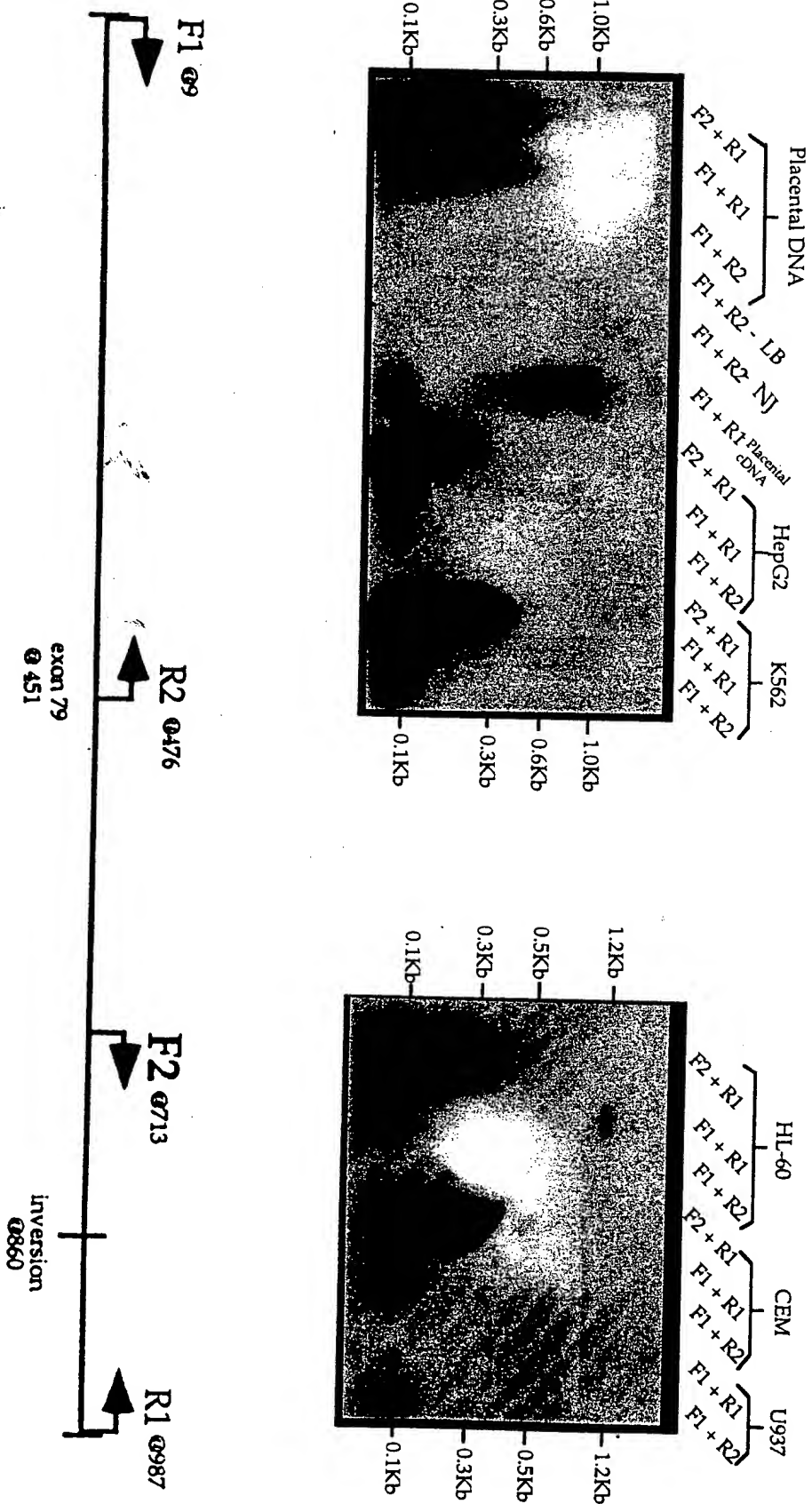


*RT-PCR primers (bold) and Southern Blotting probes are indicated by the arrows.

Figure 17B

The Apo-dystrophin cDNA

Figure 17C





RECEIVED

SEP 18 2002

TECH CENTER 1600/2900

A.

12/23bp spacer
CACAGTG-----ACAAAAACC
heptamer nonamer

Figure 18A

B.

inversion breakpoint₁

	11640	11650	11660	11670	11680
dystrophin	T TTATAACAGT TATAAAGAAA GA [^] TTGTAAAC TAAAGTGTGC				
	A AATATTGTCA ATATTTCTTT CT [^] AACATTG ATTTCACACG				
apo-4 cDNA	840	850		870	
[138]	T TTATAACAGT TATAAAGAAA GA [^] TTaTAAAg gAAAAaGaaa>				
	^ ^^^^^^^^^^ ^^^^^^^^^^ ^^ ^^v^^^v v^^^vv^vvv				
dystrophin	T TTATAACAGT TATAAAGAAA GA [^] TTGTAAAC TAAAGTGTGC				

	11690	11700	11710	11720	11730
dystrophin	TTTATAAAAA AAAGTTGTTT ATAAAAACCC CTAAAAACAA AACAAACACA				
	AAATATTTTT TTTCAACAAA TATTTTTGGG GATTTTTGTT TTGTTTGTGT				
apo-4 cDNA	880	890	900	910	920
[138]	aTaAaAtggA cAGTgGTgA ATgtgAAcTc aggtgtgCac AAttAtCagg>				
	v^v^v^vvv^ v^^^v^^v ^^vv^^^v vvvvvv^^v ^^vv^v^^v				
dystrophin	TTTATAAAAA AAAGTTGTTT ATAAAAACCC CTAAAAACAA AACAAACACA				

	11740	11750
dystrophin	CACACACACA CATAACACACA	
	GTGTGTGTGT GTATGTGTGT	
apo-4 cDNA	940	950
[138]	aACAC-CcCA -AaAC-CAaA>	
	v^^^^ ^v^^ ^v^^ ^^v^	
dystrophin	CACACACACA CATAACACACA	

Figure 18B (SEQ ID NO: 57)



	13130	13140	13150	13160	13170
	*	*	*	*	*
dystrophin	AATTAGCTTT	TGGAGAGTGG	GTTTTGTCCA	TTATTAATAA	TTAATTAATT
	TTAATCGAAA	ACCTCTCACC	CAAAACAGGT	AATAATTATT	AATTAATTAA

inversion breakpoint₂

	13180	13190	13200	13210	13220
	*	*	*	*	*
dystrophin	AACATCAAAC	ACGGCTTCTC	ATGCTATTTC	TACCTCACTT	TGGTTTTGGG
	TTGTAGTTTG	TGCCGAAGAG	TACGATAAAG	ATGGAGTGAA	ACCAAAACCC

	980	970	960	950	940
apo-4	<AACATCAAAC	ACGGCTTCTC	ATGCTATTTC	TACCTCACTT	TGGTTTTGGG
	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^
dystrophin	AACATCAAAC	ACGGCTTCTC	ATGCTATTTC	TACCTCACTT	TGGTTTTGGG

	13230	13240	13250	13260	13270
	*	*	*	*	*
dystrophin	GTGTTCTGA	TAATTGTGCA	CACCTGAGTT	CACAGCTTCA	CCACTTGTCC
	CACAAGGACT	ATTAACACGT	GTGGACTCAA	GTGTCGAAGT	GGTGAACAGG

	930	920	910	900	890
apo-4	<GTGTTCTGA	TAATTGTGCA	CACCTGAGTT	CACAGCTTCA	CCACTTGTCC
	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^
dystrophin	GTGTTCTGA	TAATTGTGCA	CACCTGAGTT	CACAGCTTCA	CCACTTGTCC

	13280	13290	13300	13310	13320
	*	*	*	*	*
dystrophin	ATTGCGTTAT	TTTCTTTTTT	CTTTATAATT	CTTCTTTTTT	CCTTCATAAT
	TAACGCAATA	AAAGAAAAAG	GAAATATTAA	GAAAGAAAAA	GGAAGTATTA

inversion breakpoint₃

	880	870	860	850	840
apo-4	<ATTGCGTTAT	TTTCTTTTTT	CTTTATAATT	CTTCTTTTaT	aacTgtTAta
	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^	^^^^^^^^^^
dystrophin	ATTGCGTTAT	TTTCTTTTTT	CTTTATAATT	CTTCTTTTTT	CCTTCATAAT

Figure 18C (SEQ ID NO: 58)

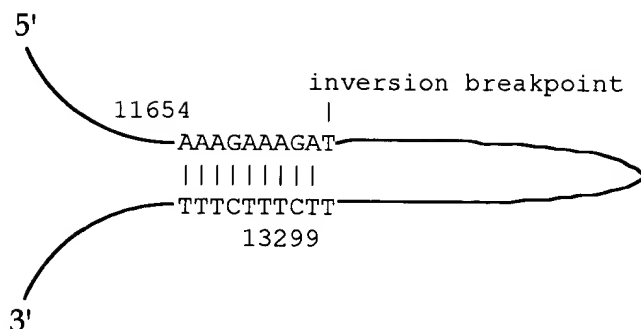


Figure 18D (Sequence included in SEQ ID NO: 57)

inversion @ 860

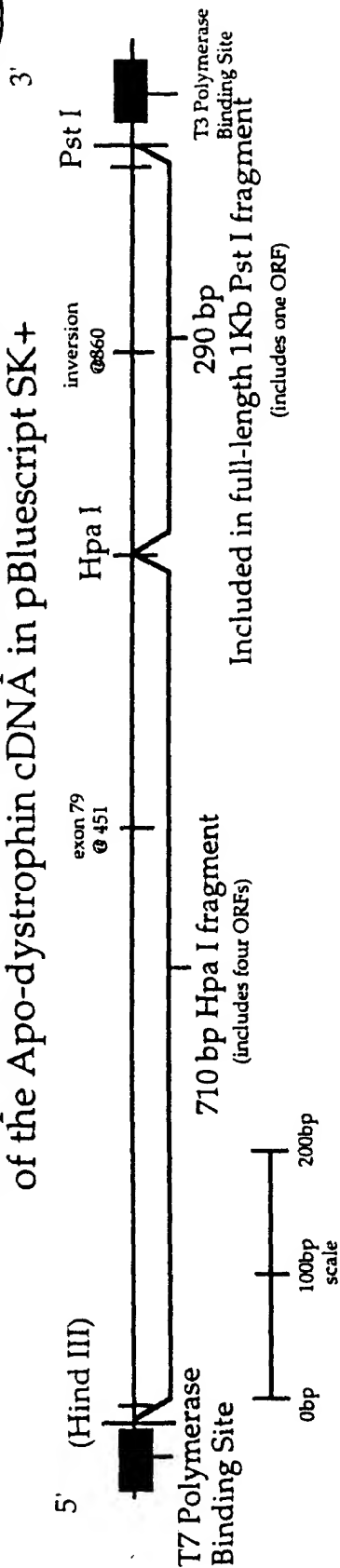
841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900

ATTGTCAATATTTCTTTCTTTAATATTTCCTTTTCTTTTATTGCGTTACCTG TTCACCAC

Figure 19 (Sequence included in SEQ ID NO: 2)



Description of *In Vitro* Transcription and Translation of the Apo-dystrophin cDNA in pBluescript SK+



Linearize plasmid with either HpaI (truncated) or Pst I (full length). Gene Clean and incubate with T7 polymerase and dNTPs to produce RNA *in vitro*.

Incubate RNA with Wheat Germ Extract or Rabbit Reticulocyte Lysates to produce *in vitro* translation

Separate translation products by SDS-PAGE. Fix, Amplify and Dry Gel. Perform Autoradiography

Figure 20

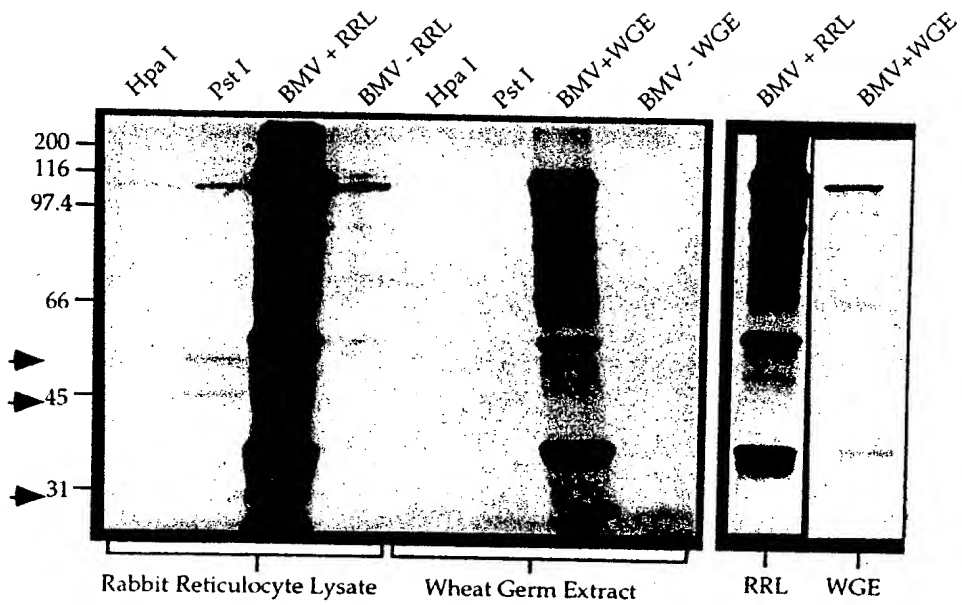


Figure 20A

Figure 20B

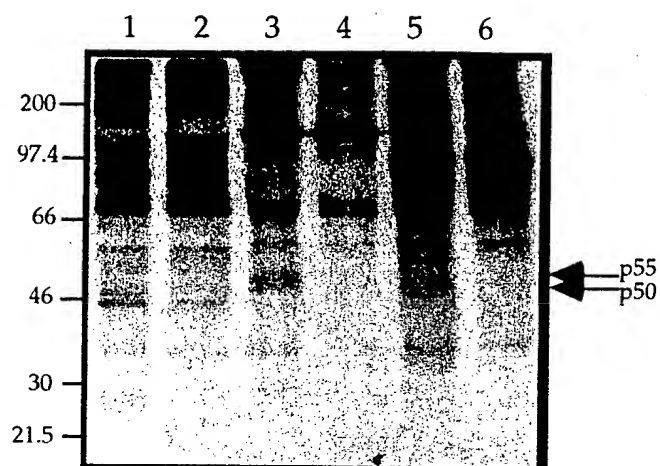


Figure 21

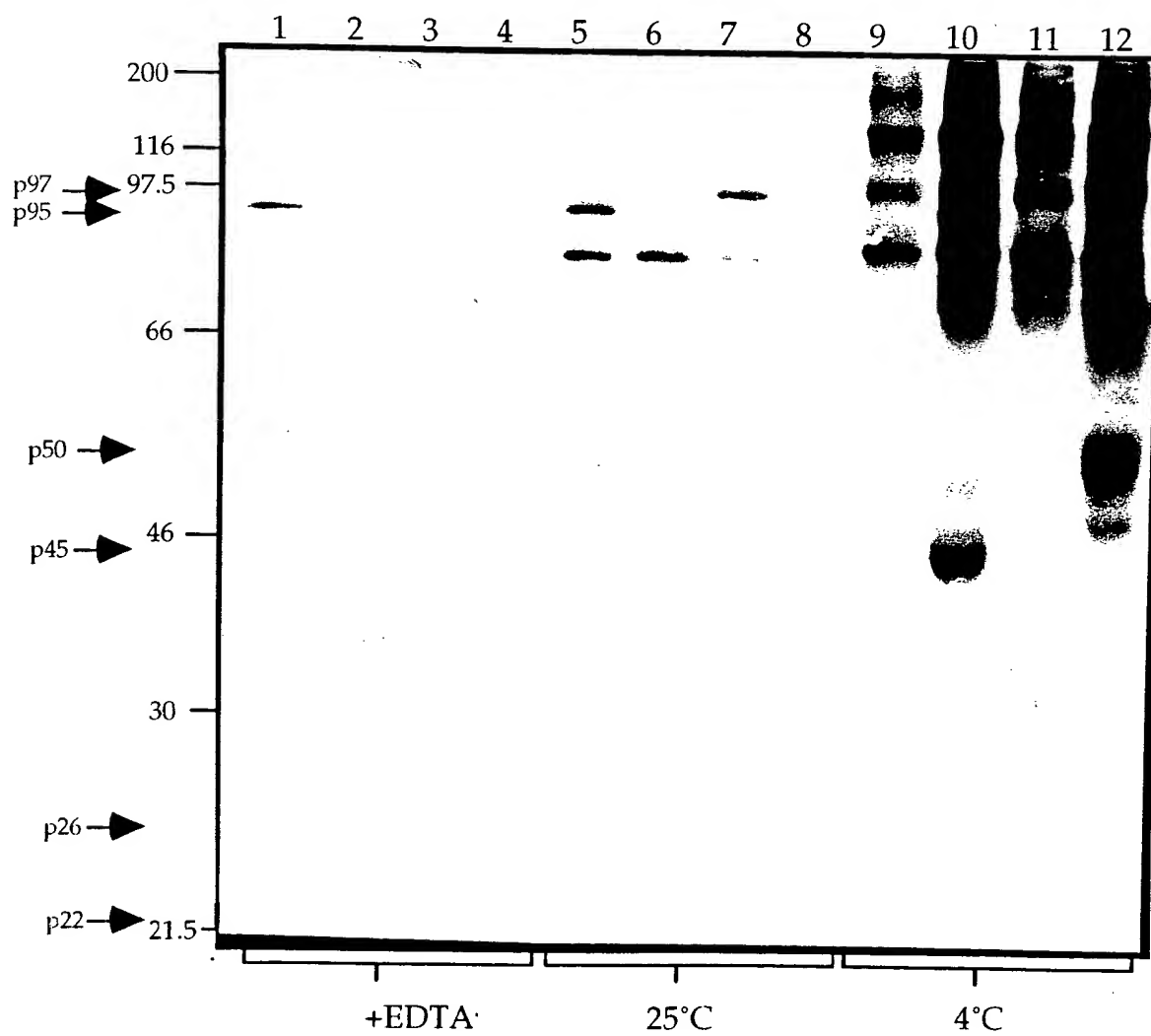


Figure 22

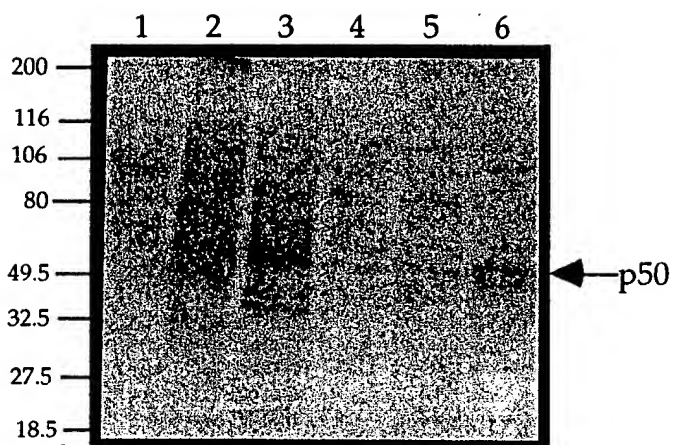


Figure 23

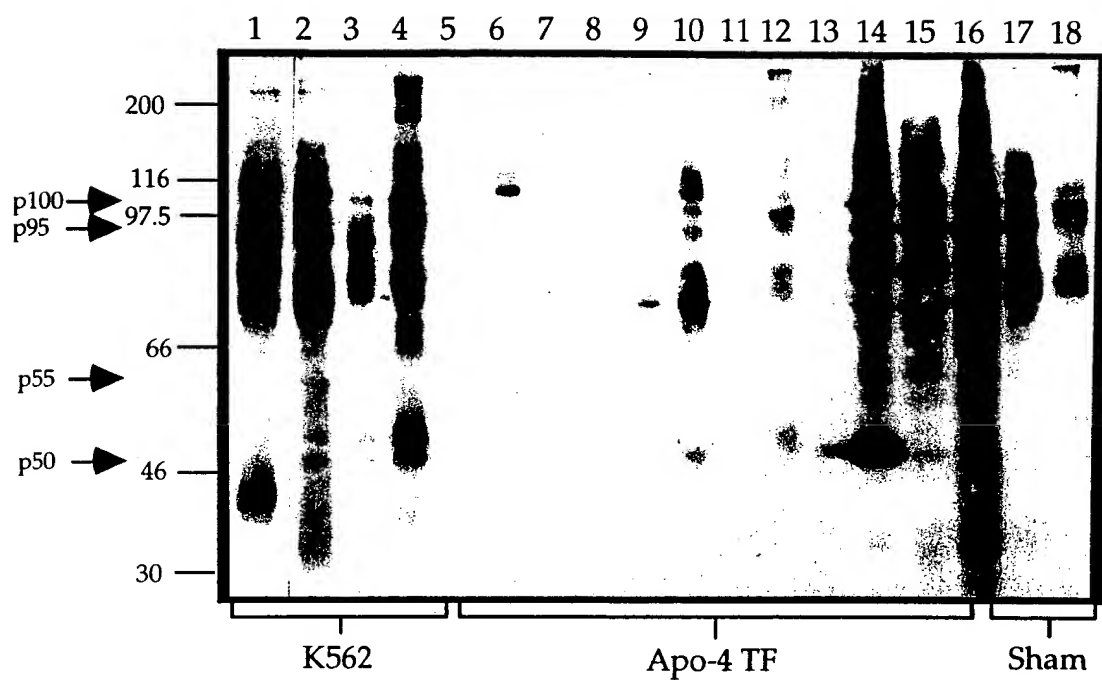


Figure 24

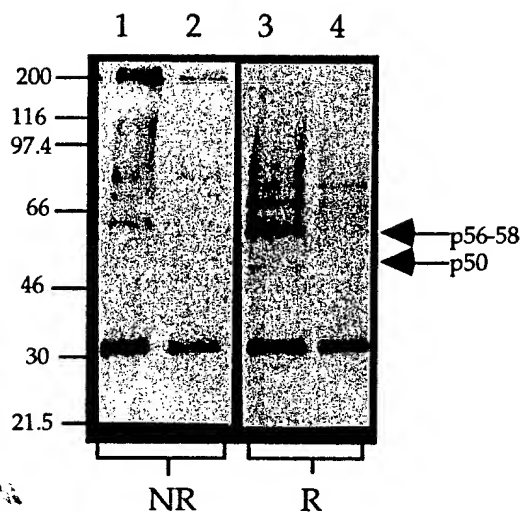


Figure 25A

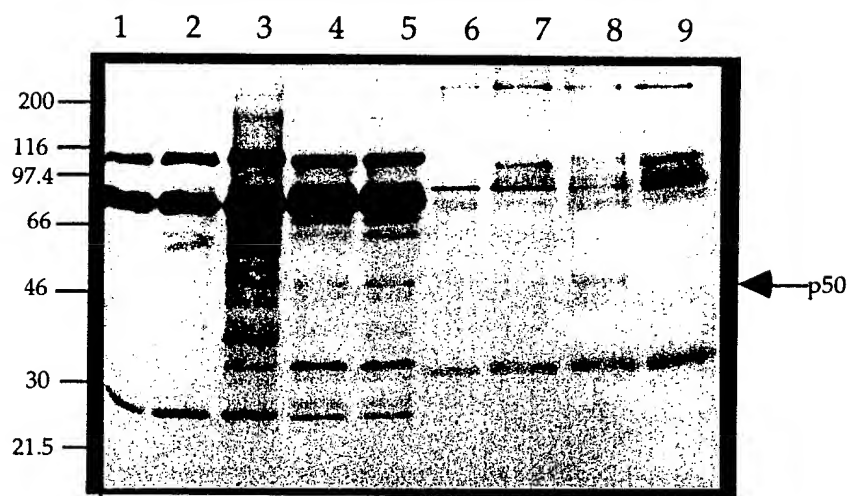


Figure 25B

H2 starting at second methionine - 321 bp, predicted weight = 17.4Kd + 1 N-glycosylation site + 20.4 Kd.

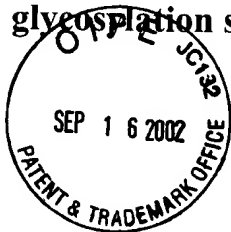


Figure 26A

Splice sites for peptide

MYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPDSRMVFIIFI
 QQRGLDSKSLQEINLYFCEGFYTSMQLYKKVIRKLHKITQWTRTPQNQSEV
 EIA (SEQ ID NO: 61)

Figure 26B

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@88 bp	78.3	@74-180	106 bp	79.1	@181-529	349 bp
	79.1	@530-654	125 bp	79.4	@655-720	66 bp
	79.4	@721-769	49 bp	79.55	@770-875	105 bp
	79.55	@876-893	18 bp	79.75	@894-932	39 bp
	79.85	@933-966	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 12- 32 1.8833

Figure 26C

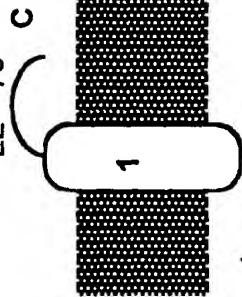


Predicted TM structure

> : Too long to be significative
< : Too short to be significative
LI: Loop length
KR : Number of Lys and Arg

KR Diff : Positive charge difference
CE : Net charge energy
CE Diff : Net charge difference
CH Diff : Charge difference over N-term segments

CE = -0.54
KR =>
LL = 75



LL = 11
KR = 1
CE = <

KR Diff = 1 CYTOPLASM
CH Diff = -3 OUTSIDE
CE Diff = 0.54 OUTSIDE

Structure no. 1

Figure 26D



A readthrough apo-4S product using the second available methionine

The Apo-4S peptide sequence

P1 Begin TM1(R)

+30 | P2

MYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHF**SMKTP**
VAR**SNIKLIL** 80

TNNVKWLHKK **GFASSWKL****VK** **NQTLCTPSM QLLCCLHP****EM**
GNDFPNGKET 130

P3

ERCYLSAPFV KSVFLSLCFP **GHN****VGSLFHM** **ADDLGRAMES**
LVSVMTDEEG 180

AEKMFYNSRF **PHGFYNIHTT** **KRIRQKEFTR** **NKSIFL****RRVV VLYCRFQKFL**
230

LLLLFCKQWQ **VLHVYAIVQK** **SYKKT****TTCKIL IAKKL****LAISLY**
GTHFGLFKNL 280

KQLKRKNYKG **KRKKRNGQVV** **KLRTQVCTII** **RNTPKPKRGR**
NSMRSRVRCK 330

LI 332 (302aa in predicted polypeptide) (SEQ ID NO: 56)

Figure 27A



Candidate membrane-spanning segments:

Certain	1	41-61	1.9073
Putative	2	101-121	0.8052
Certain	3	132-152	1.2552
Putative	4	217-237	1.1833
Putative	5	254-274	0.9240

Transmembrane segments included in structure No. 8: 1 2 3 4 5

Loop lengths: 11 39 10 64 16 58; K+R profile: 1 2 5 (9 > 22)

K+R difference: -23: -> Orientation: **N-out**

Charge-difference over N-terminal Membr. segs. (± 15 residues): -4
-> Orientation: **N-out**

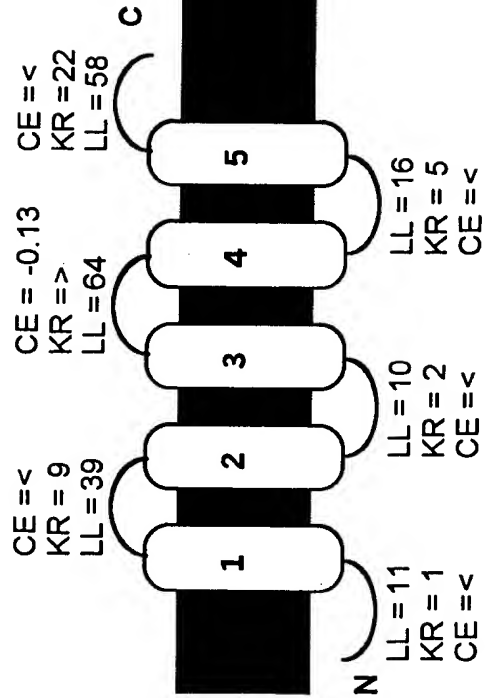
CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 <

CYT-EXT difference: 0.13: -> Orientation: **N-out**

Figure 27B



> : Too long to be significant
 < : Too short to be significant
 LL : Loop length
 KR : Number of Lys and Arg
 KR Diff : Positive charge difference
 CE : Net charge energy
 CE Diff : Net charge difference
 CH Diff : Charge difference over N-term segments



KR Diff = -23 OUTSIDE
 CH Diff = -.4 OUTSIDE
 CE Diff = 0.13 OUTSIDE

Structure no. 8

Figure 27C

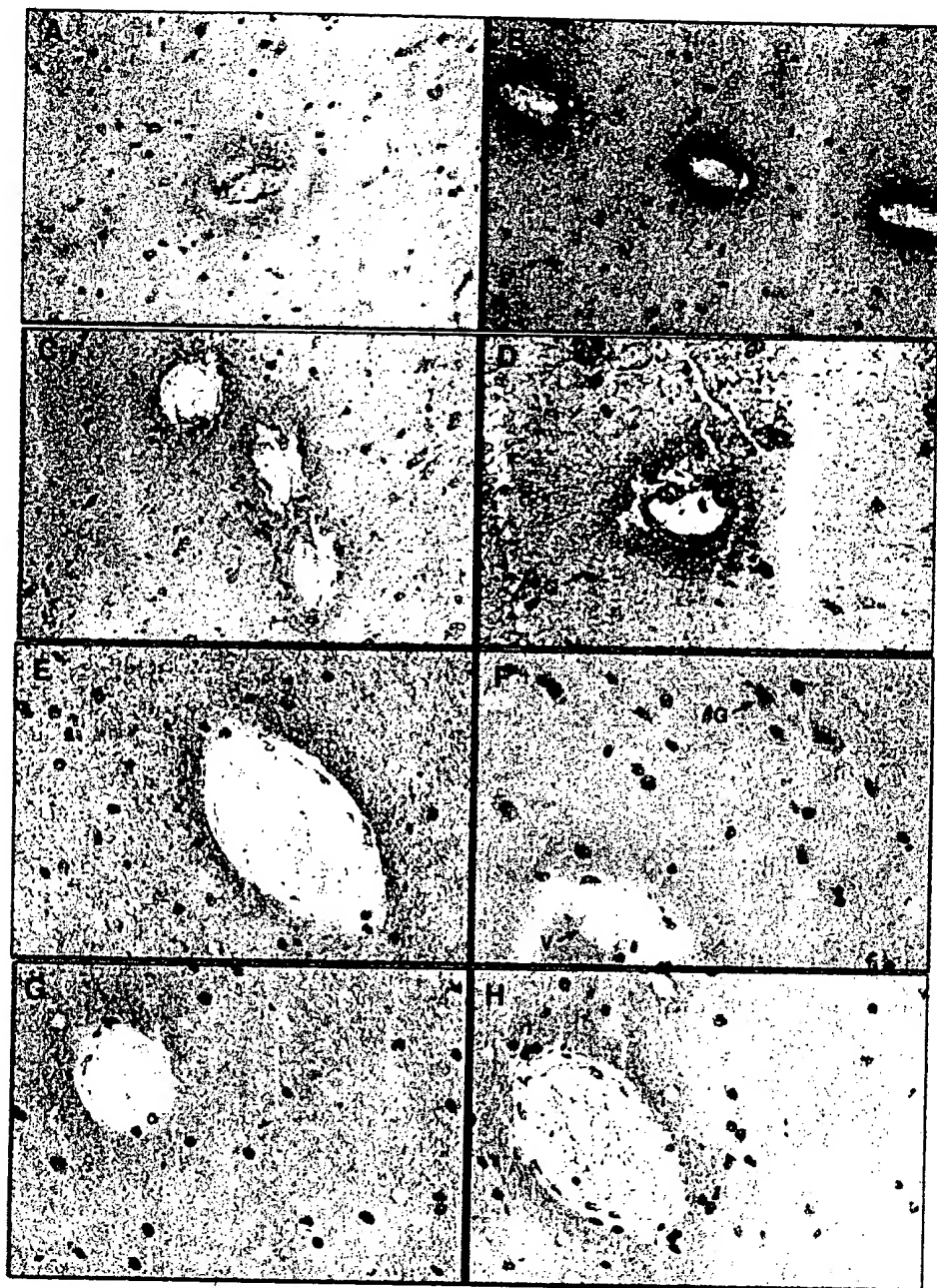


Figure 28

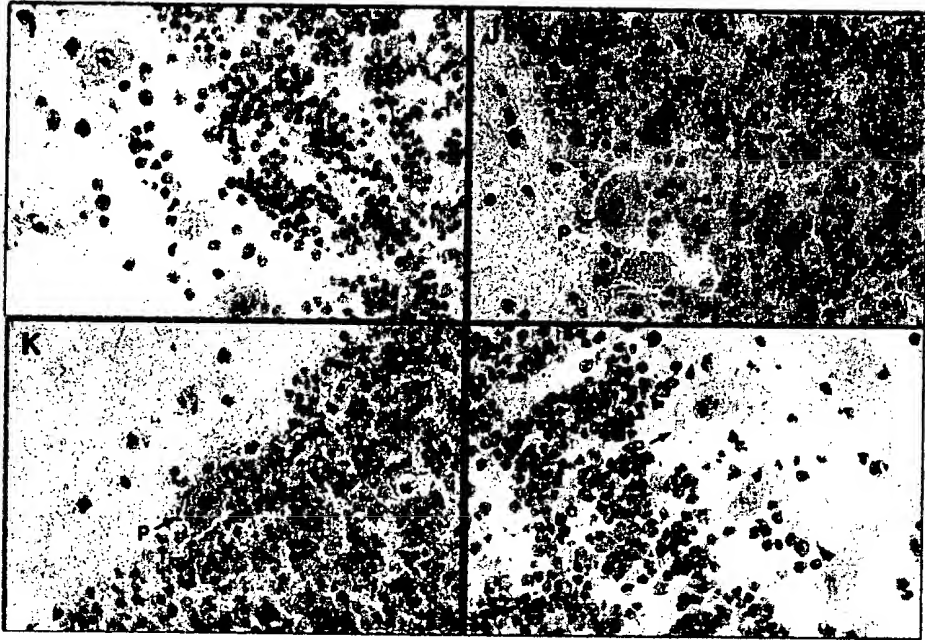


Figure 28 (cont'd)

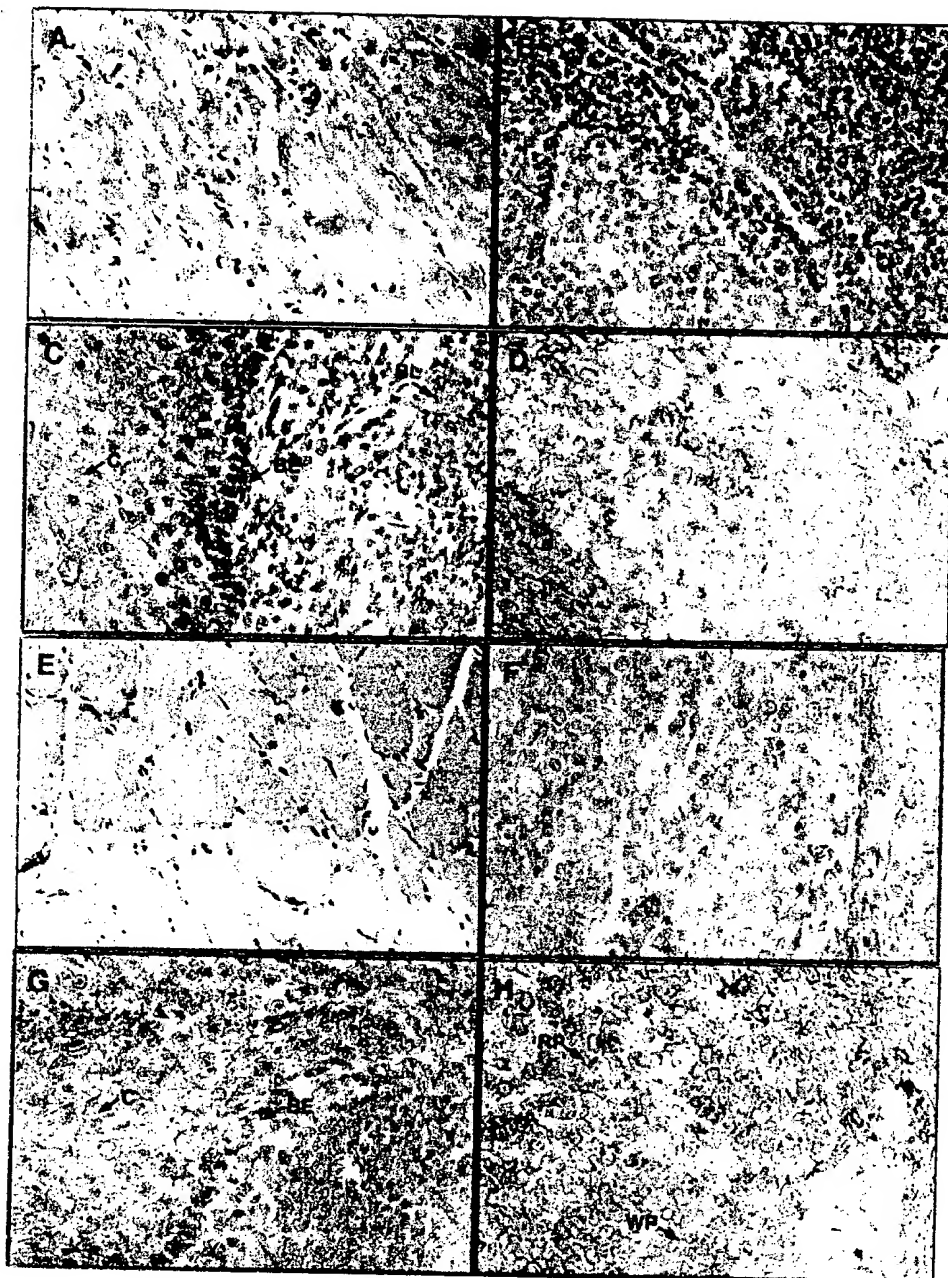


Figure 29

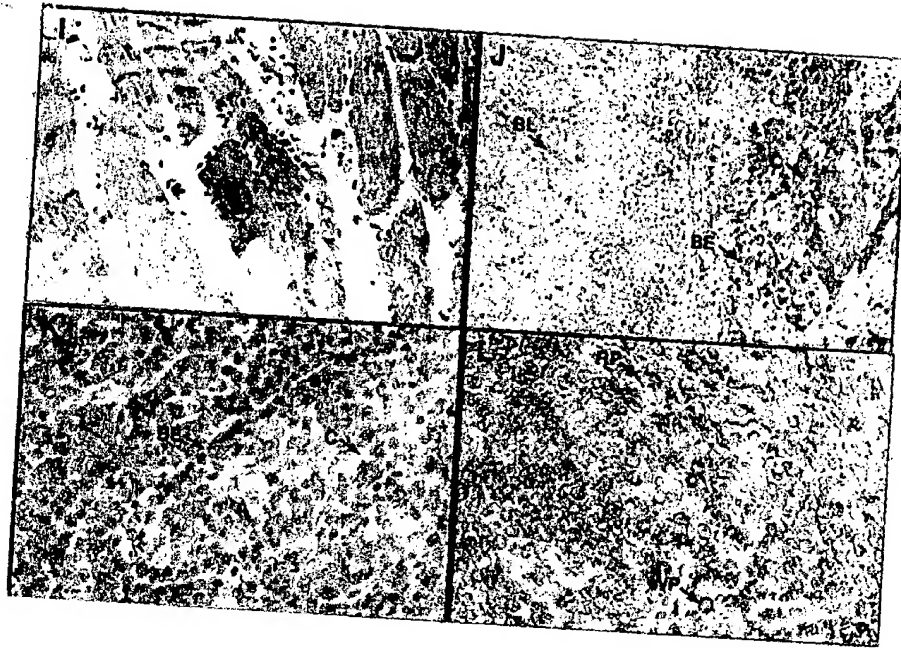


Figure 29 (cont'd)

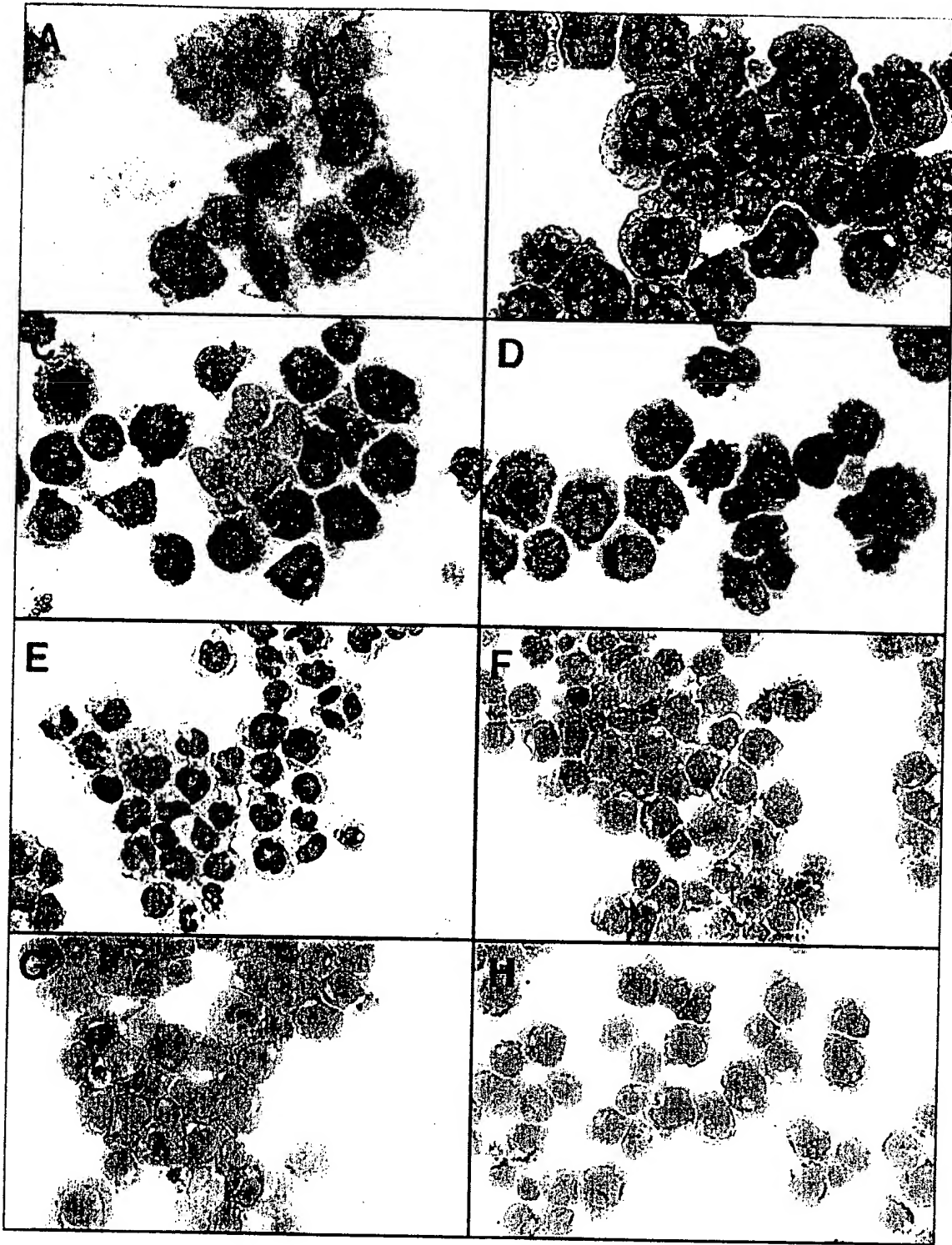


Figure 30



RECEIVED
SEP 18 2002
TECH CENTER 1600/2900

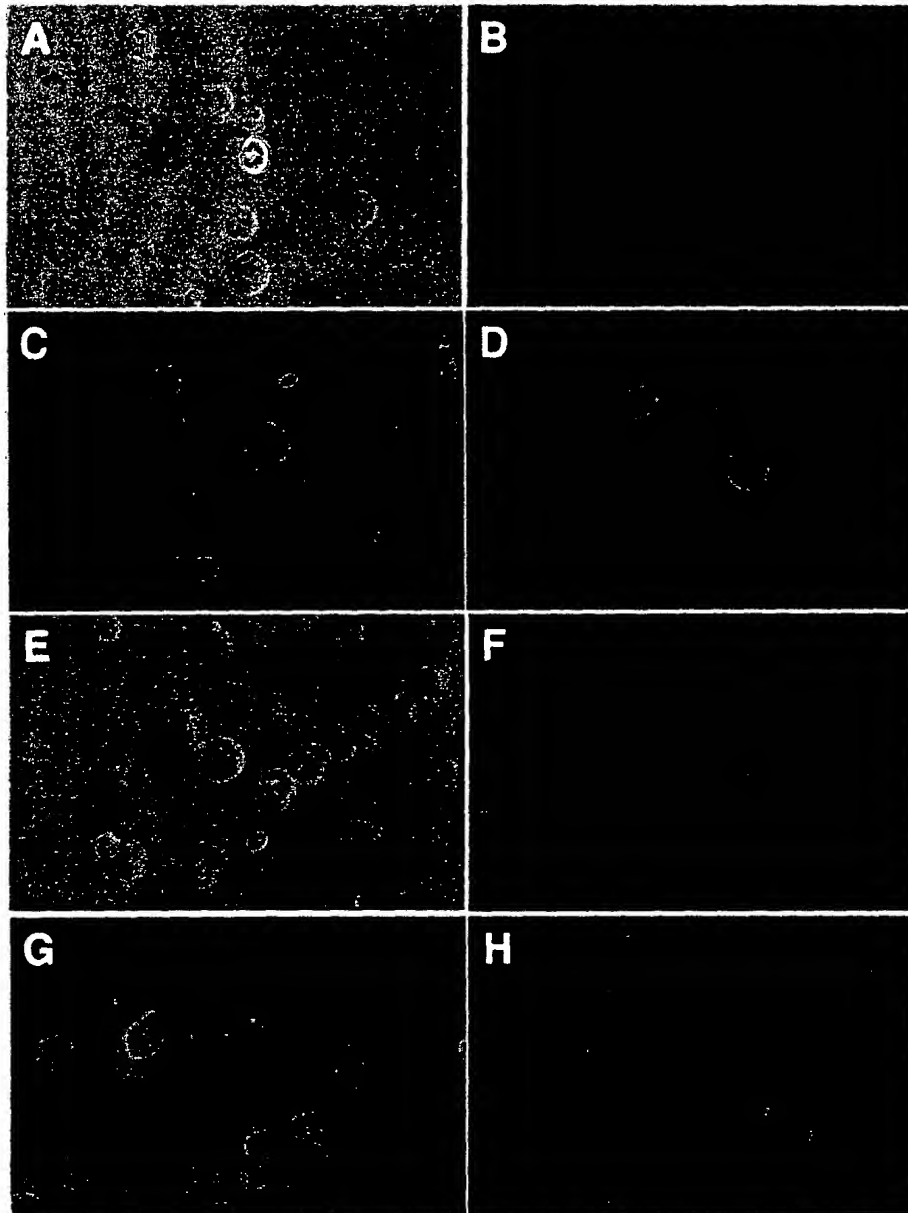


Figure 31

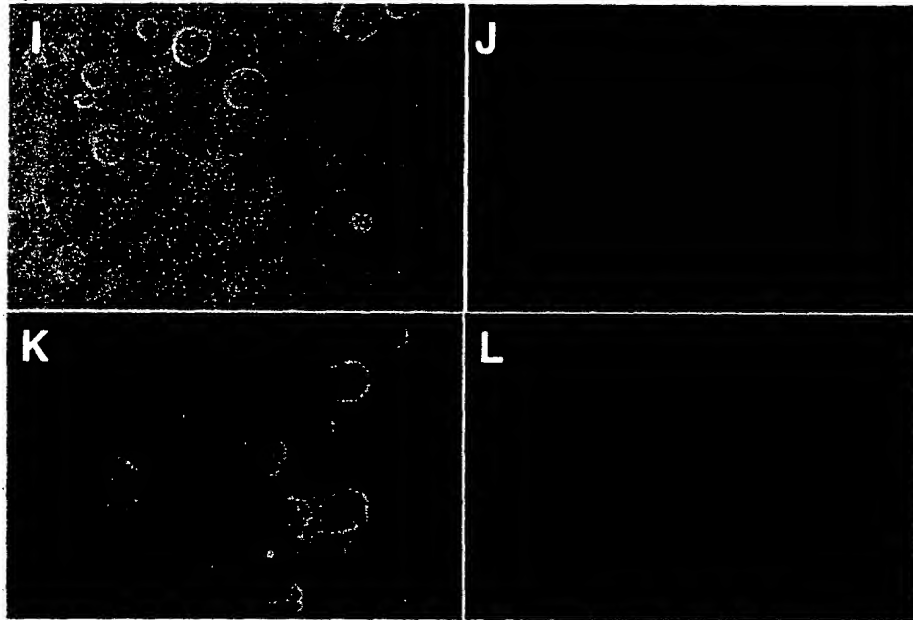


Figure 31 (cont'd)

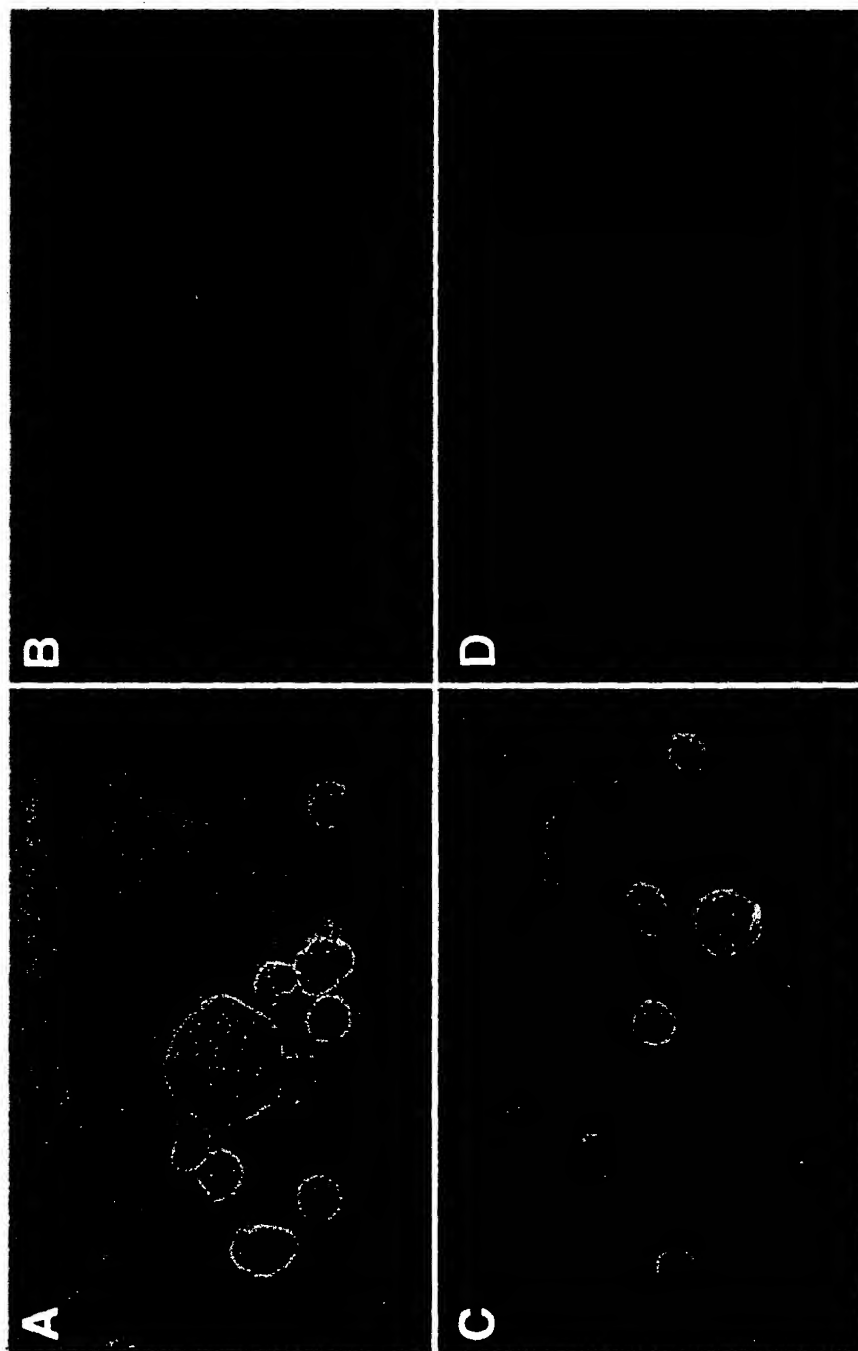


Figure 32



RECEIVED
SEP 18 2002
TECH CENTER 1600/2900

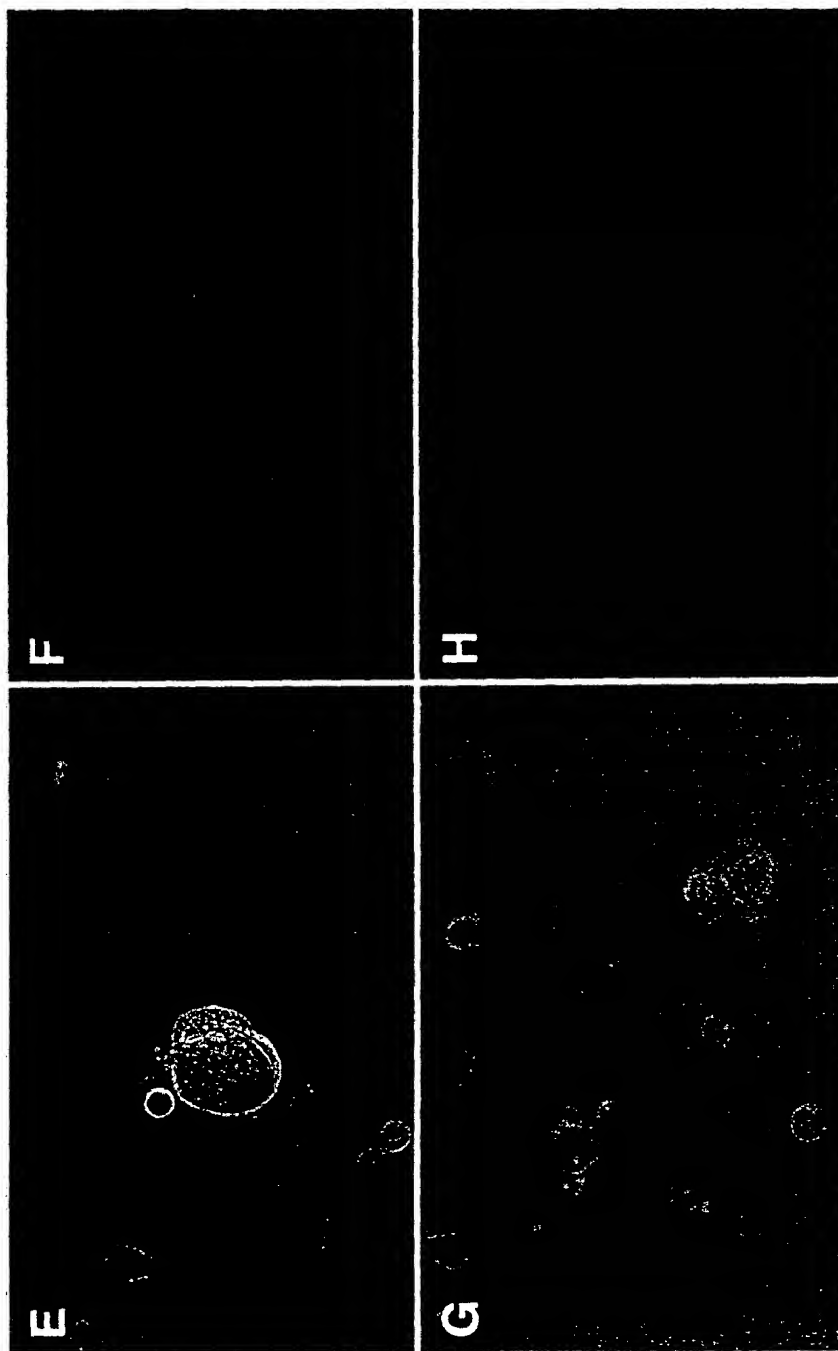


Figure 32 (cont'd)

SEP 18 2002

TECH CENTER 1600/2900

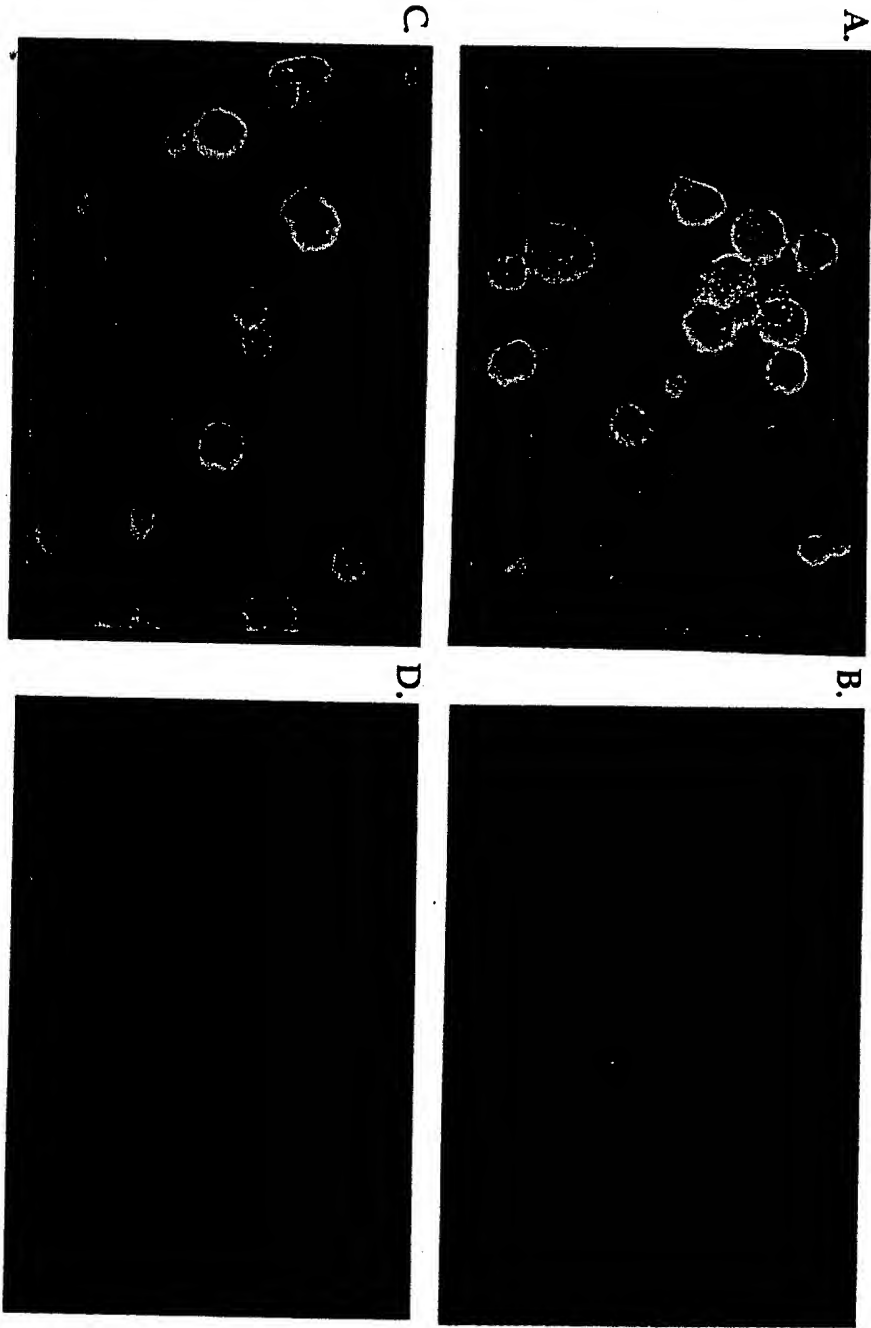


Figure 33



**Additional Oligonucleotide primers used for apo-dystrophin-4
southern blotting and sequencing**

FORWARD

GTT CGT TAA TAC AAG TAG	F2.3(@28)	(SEQ ID NO 15)
GCC AAG GTG GAA AAG ATG	F2.2(@73)	(SEQ ID NO 16)
CCA GTA GCC TGA TCC AAC	F3.2(@208)	(SEQ ID NO 17)
GGC TTC ATT AAT AAG	F3.1(@257)	(SEQ ID NO 18)
GGC AAA GAA ACA GAG TG	F4.2(@379)	(SEQ ID NO 19)
CAG GAC ACA ATG TAG GA	F4.1(@449)	(SEQ ID NO 20)
GTT ATA AAG AAA GAA TTA TAA AG	FJn(@846)	(SEQ ID NO 21)
GAA AAT AAC GCA ATG GAC	F5.1(@875)	(SEQ ID NO 22)

REVERSE

GAT GGG ATA CAT CTT TTC C	R6.1(@99)	(SEQ ID NO 23)
CAA GCT ACA TTC AGG TTC CC	F2.2R(@188)	(SEQ ID NO 24)
GGA CTC CAT CGC TCT GCC	R4.1(@510)	(SEQ ID NO 25)
GAC TTA GAA ACT ACT G	R3.4(@694)	(SEQ ID NO 26)
ATA GAC GTG TAA AAC CTG C	R2.1(@735)	(SEQ ID NO 27)
AAC TGT TAT AAA TTT TTA	RSP2(@848)	(SEQ ID NO 28)
CTT TTT CCT TTA TAA TTC TTT C	R2.3o(@875)	(SEQ ID NO 29)

Figure 34



An Additional Splice Product Predicted From The Apo-4 Gene

A second potential theoretical splice product which retains exon 78.3 is shown below.

H2 p1-124 spliced product =351 bp, 117 amino acids + 10 from vector + 1 N-glycosylation site; predicted weight = 21.9 Kd

Figure 35A

Peptide Generated

MFVNTTKVEKMYPIMEYSCSDRNLVLIIYGILLIYIIGNLNMKKEQNKCFCTTPDSRMVFII
FIQQRGLDSKSLQEINLYFCEGFYTSMLYKKVIRKLHKITQWTRTPQNPQSEVEIA (117
amino acids) (SEQ ID NO 30)

Figure 35B

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@26 bp	78.1	@16-41	26 bp	78.3	@42-74	33 bp
	78.3	@75-181	106 bp	79.1	@182-530	349 bp
	79.1	@531-655	125 bp	79.4	@656-721	66 bp
	79.4	@722-770	49 bp	79.55	@771-876	105 bp
	79.55	@877-894	18 bp	79.75	@895-933	39 bp
	79.85	@934- 967	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 22- 42 1.8833

Figure 35C

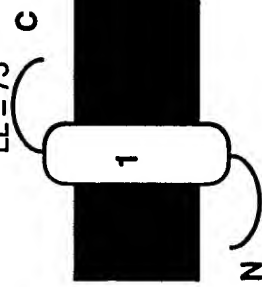


D. Predicted TM structure

- > : Too long to be significative
- < : Too short to be significative
- LI : Loop length
- KR : Number of Lys and Arg

- KR Diff : Positive charge difference
- CE : Net charge energy
- CE Diff : Net charge difference
- CH Diff : Charge difference over N-term segments

CE = -0.54
KR = >
LL = 75



LL = 21
KR = 3
CE = <

KR Diff = 3
CH Diff = -2
CE Diff = 0.54

CYTOPLASM
OUTSIDE
OUTSIDE

Structure no. 1

Figure 35D



Nucleic Acid Subsequence Sites Identified In Apo-4

<u>Motif</u>	<u>Position</u>	<u>Significance</u>
CpG	-7, (+28, +106)	DNA methylation site
CAAT	-132, (+127, +131)	Binding of CAAT factors
TATAAT (5/6)	-120, -114, (+10)	TFIID Binding site
TATA	-154	Binds RNA polymerase II and TFIID
CCATTCA	-162, -131	Cap Site I
TATCAGT	+12, (+25)	Cap Site II
TGGCTGCAAGCCCAA (10/14) (SEQ ID NO: 32)	-57, (+41)	Binds CTF/NF-I protein
GTGATGG	-140, -4, +11, +32	Eucaryotic Transcription Initiation Site

Figure 36



Top Pred predicts 4-5 transmembrane domains for a full-length apo-4F product in which all the stop codons are suppressed.

Protein sequence and position of predicted TM domains

Begin TM₁(R)

P1 | P2

MFVNTSREKV INQSLIAKVE KMYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMK60

TPVARSNIKL ILTNNVKWLH KKGFASSWKL VKNQTLCTP SMQLLCLHP EMGNDFPNGK 120

P3

ETERCYLSAP FVKSVFLSLC FPGHNVGSLF HMADDLGRAM ESLVSMTDE EGAEKMFYNS180

RFPHGIFYNIH TTKRIRQKEF TRNKSIFLRR VVLYCRFQK FLSLLLFCKQ WQVLHVYAIV 240

QKSYKKTCK ILIAKKLAIS LYGTHFGLFK NLKQLKRKNY KGKRKKRNGQ VVKLRTQVCT 300

IIRNTPKPKR GRNSMRSRVR CKLI (324 amino acids) (SEQ ID NO 31)

Hydrophobicity Scale KD

Figure 37A

Apo-4F : Candidate membrane-spanning segments:

Certain	1	33- 53	1.9073
Putative	2	93- 113	0.8052
Certain	3	124- 144	1.2552
Putative	4	209- 229	1.1833
Putative	5	246- 266	0.9240

I. Transmembrane segments included in structure 8: 1 2 3 4 5; Loop lengths: 32 39 10 64 16 58

Figure 37B



K+R difference: -19; -> Orientation: **N-out**; Charge-difference over N-terminal Membr. segs.

(±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < < < < -0.13 <

CYT-EXT difference: 0.13

-> Orientation: **N-out**

II. Transmembrane segments included in structure 7: **1 3 4 5**; Loop lengths: 32 70 64 16
58

K+R profile: 5 > 22 > 5; K+R difference: 22 -> Orientation: **N-in**

Charge-difference over N-terminal Membr. segs. (±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 < -0.26 <

CYT-EXT difference: 0.13; -> Orientation: **N-out**

Figure 37B (cont'd)



C. TopPred predicts a cytoplasmic N-terminus for four TM domains

> : Too long to be significative
 < : Too short to be significative
 LI : Loop length
 KR : Number of Lys and Arg

KR Diff : Positive charge difference
 CE : Net charge energy
 CE Diff : Net charge difference
 CH Diff : Charge difference over N-term segments

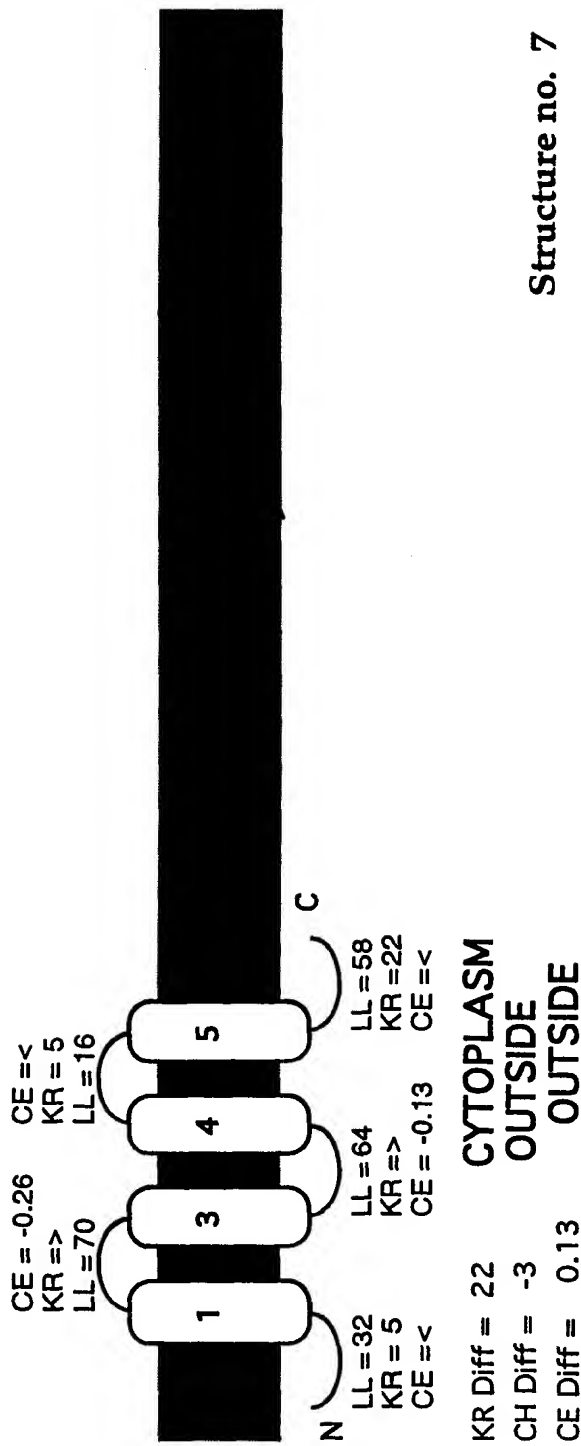


Figure 37C

Basic Features of a Transposon or Retrovirus

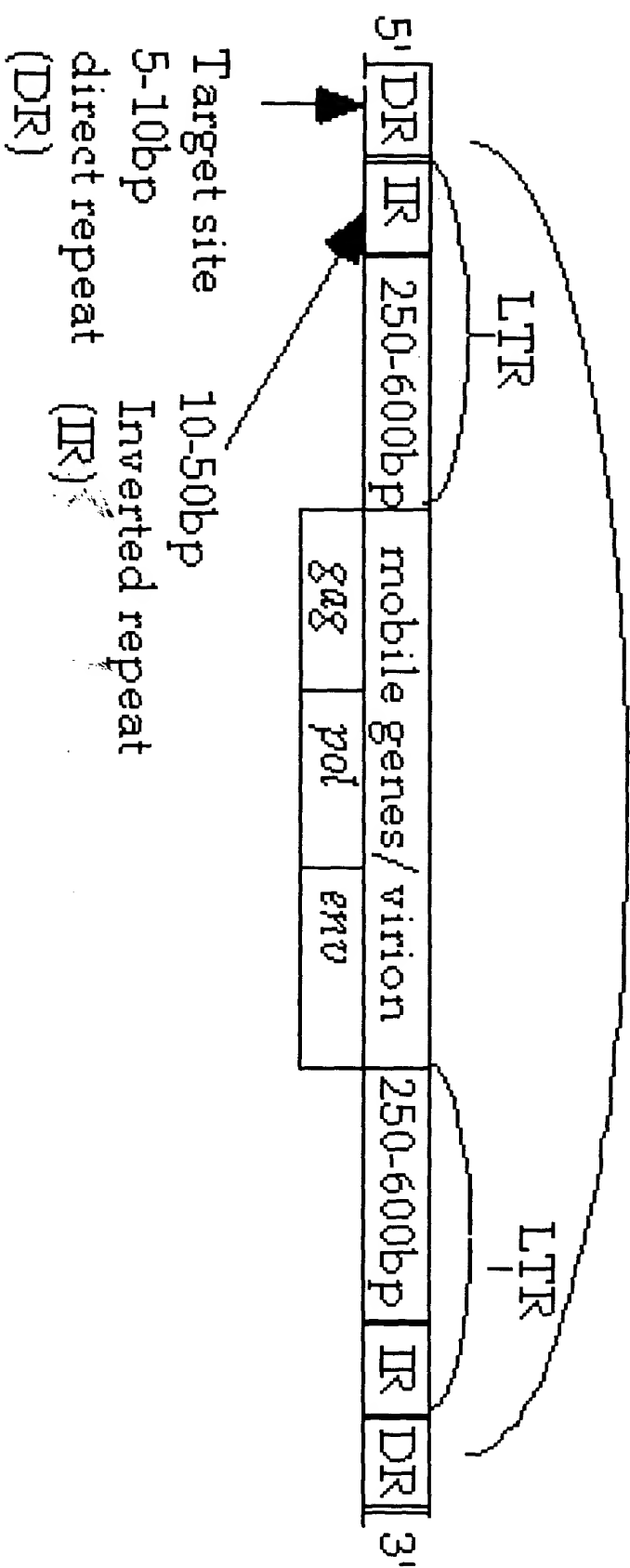


Figure 38A



Structure of the apo-4 inversion element before rearrangement

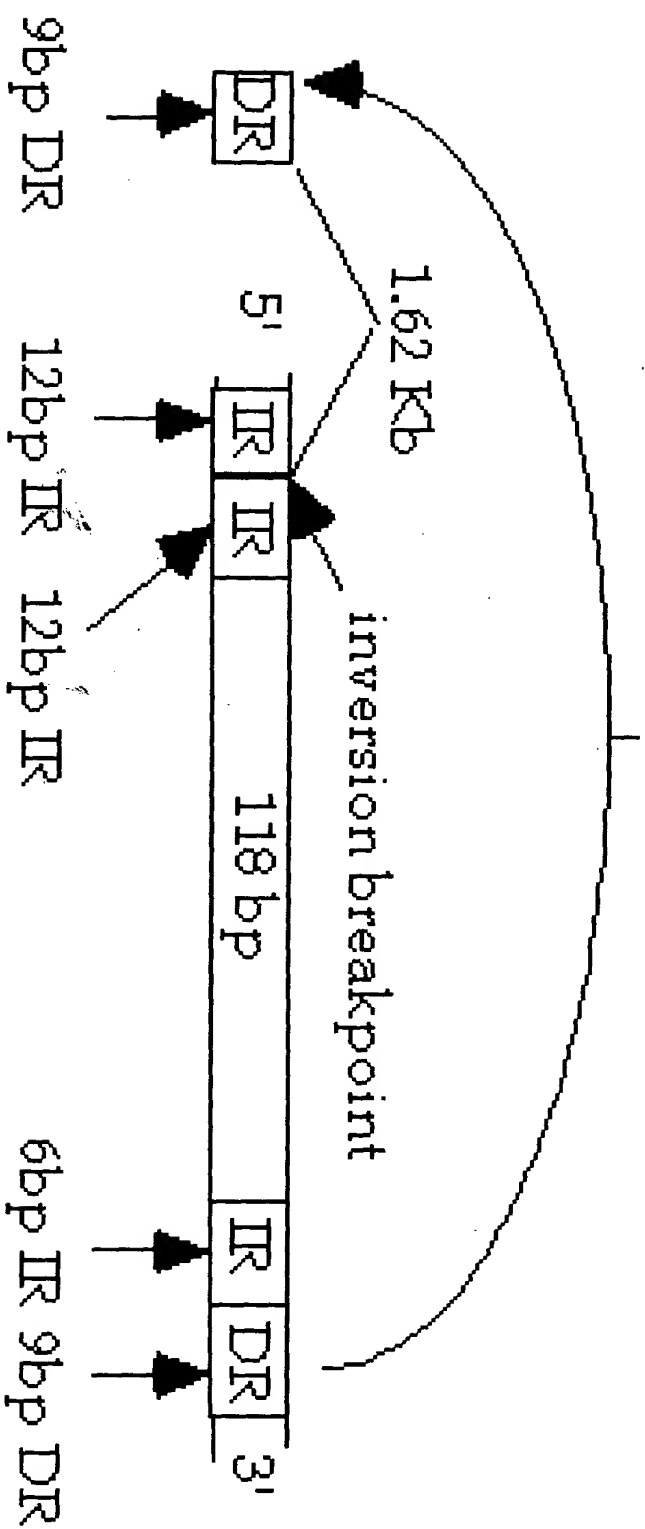
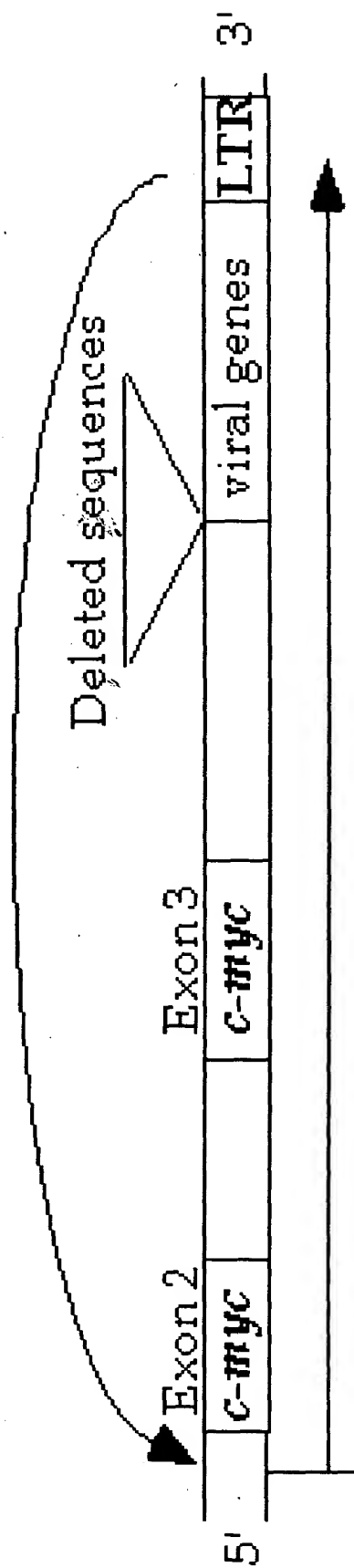


Figure 38B





RNA transcript is promoted from cell sequences but enhanced and terminated by viral sequences.

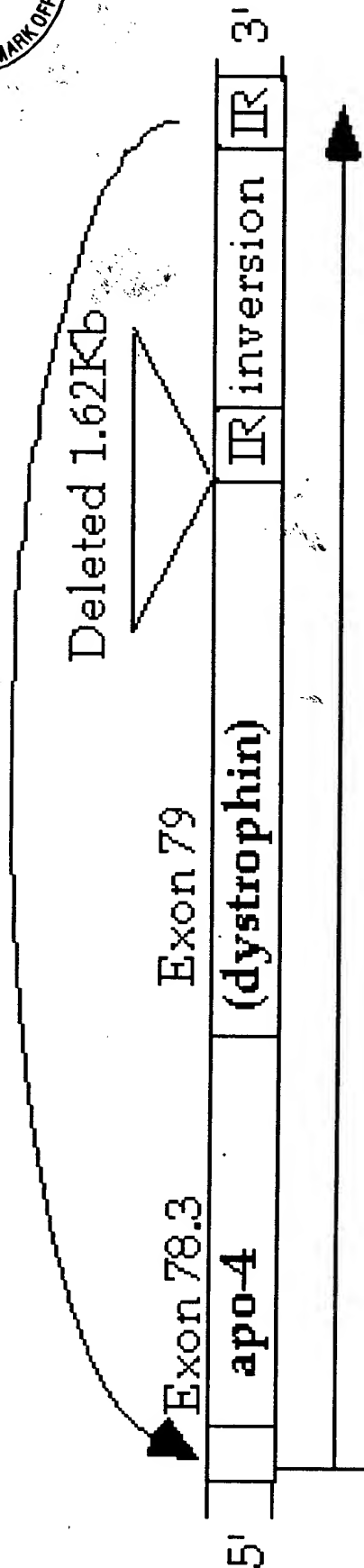
Figure 39A



RECEIVED

SEP 18 2002

TECH CENTER 1600/2900



RNA transcript is promoted from cell sequences but enhanced and terminated by inversion sequences which may also activate suppressor tRNAs or reverse transcriptase activity to prevent the recognition of stop codons. Inverted repeats (IR) are present at both ends of the inversion, as they are in retroviruses and transposable elements.

Figure 39B